(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 July 2002 (18.07.2002)

PCT

(10) International Publication Number
WO 02/056025 A2

(51) International Patent Classification[7]: G01N 33/68

(21) International Application Number: PCT/IB01/02831

(22) International Filing Date:
30 November 2001 (30.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/727,812 30 November 2000 (30.11.2000) US

(71) Applicant *(for all designated States except US)*: **INTEGRATIVE PROTEOMICS, INC.** [CA/CA]; 100 College Street, Suite 521, Toronto, Ontario M5G 1L5 (CA).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **AWREY, Donald, E.** [CA/CA]; 2211 Stir Crescent, Mississauga, Ontario L4Y 3V2 (CA). **GREENBLATT, Jack** [CA/CA]; 141 Highbourne Road, Toronto, Ontario M5P 2J8 (CA).

(74) Agent: **BERESKIN & PARR**; 40 King Street West, 40th Floor, Toronto, Ontario M5H 3Y2 (CA).

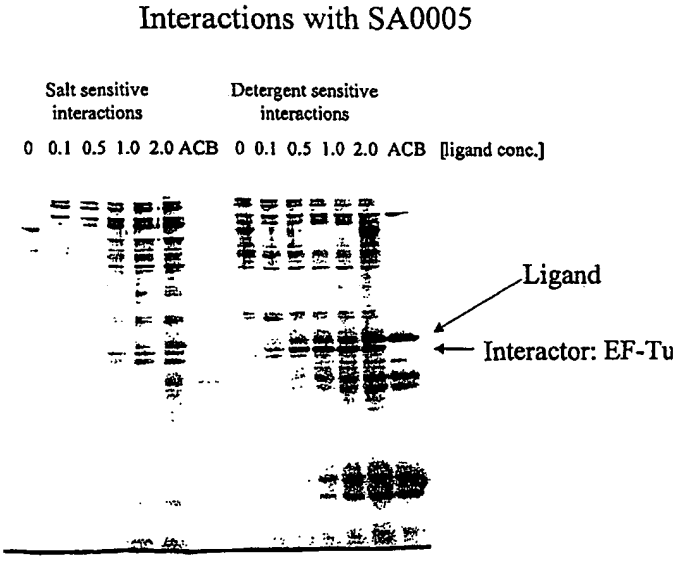
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

[Continued on next page]

(54) Title: METHODS FOR SYSTEMATIC IDENTIFICATION OF PROTEIN - PROTEIN INTERACTIONS AND OTHER PROPERTIES



(57) Abstract: The present invention relates in part to a method for identifying protein-protein interactions. The interacting proteins may be isolated by affinity chromatography and may be identified and characterized by mass spectrometry. The invention in part allows for the high throughput analysis of protein-protein interactions that lends itself to automation.

WO 02/056025 A2

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHODS FOR SYSTEMATIC IDENTIFICATION OF PROTEIN - PROTEIN INTERACTIONS AND OTHER PROPERTIES

INTRODUCTION

The genome sequencing projects are providing vast amounts of information. With the whole genome of many organisms, including humans, complete or nearing completion, the next challenge involves the characterization of the gene products. However, little is known about the functions of most proteins that the genes encode, or how these proteins interact to control cellular functions.

Protein interactions are intrinsic to virtually every cellular process. Most proteins in cells function in multi-subunit complexes of proteins created by specific protein-protein interactions. Many of the protein-protein interactions involved in cellular processes are too weak to allow co-purification of the interacting species by conventional methods from cellular extracts. The relatively weak binding is generally expected as proteins that must reversibly interact with each other in the concentrated intracellular environment will rapidly dissociate in a comparatively dilute protein mixture. As the characterization of protein-protein interactions may require the in vitro reassembly of multi-subunit protein complexes, it is important to have methods for identifying and purifying all of the interacting proteins starting with one member of a protein complex.

The two-hybrid system consists of two components, a target protein (the "bait"), fused to a DNA binding domain which binds to a specific region of DNA upstream of a reporter gene, and a protein (the "prey") fused to an activation domain which, when brought in close proximity of the reporter gene, can initiate transcription. Usually the "bait" protein is known and the "prey" protein is derived from genomic or cDNA libraries in order to isolate the interacting partner to the bait. The advantage of the two-hybrid system is that when an interactor is found the gene sequence may be determined directly. This advantage is becoming increasingly less important as the full genomic sequence of many organisms becomes available, making the identification of gene sequence from protein sequence routine. The two-hybrid system yields a very high percentage of false positives, is very labor intensive and does not easily lend itself to automation, making it a poor choice for high throughput analysis.

Protein-protein interactions have commonly been detected by antibody co-immunoprecipitation. Co-immunoprecipitation depends on the strength of a secondary protein-protein interaction, rather than on direct binding to the antibody. The technique is normally limited to relatively strong interactions with $K_d \chi 10^{-9}$ M. Additionally, it is not as sensitive as protein-affinity chromatography, because the concentration of the antigen is low.

Protein-affinity chromatography offers distinct advantages as a technique for detecting protein-protein interactions. Protein affinity chromatography allows sensitive detection of protein-protein interactions. This method can detect interactions ranging in strength from K_d 10^{-5} to 10^{-10} M. This limit is within the range of the weakest interactions likely to be physiologically relevant, which is estimated to be about 10^{-3}M. Formosa et al., Methods in Enzymology 1991, 208, 24-45. An interacting protein with a $K_d > 10^{-5}$ M may not remain bound to the column when the column is washed with buffer in order to lower the nonspecific binding of proteins from the extract to the column material.

Protein-affinity chromatography tests all proteins in an extract equally for binding to the ligand protein. Thus, extract proteins that are detected have successfully competed for the interaction with the ligand protein against the rest of the population of proteins in the extract. Additionally, interactions that are dependent on a multi-subunit complex, including the ligand protein and multiple extract proteins and/or cofactors, can be detected. Both the domains of a protein and critical residues within the protein responsible for a specific interaction can be examined for affinity to extract proteins by the use of mutant derivatives of the ligand protein.

Today, the dramatic increase in gene sequence information has far outpaced the characterization of gene products. The processes of isolation and identification of protein interactors have represented a bottleneck in the characterization of protein-protein interactions. For example, many current methods for the isolation and identification of protein interactors are performed on a protein-by-protein basis with relatively low throughput.

In part, the present invention addresses some of the concerns identified above. For example, in certain embodiments, a method of the present invention provides a process for the analysis of protein-protein interactions, which may be operated in a high throughput fashion.

SUMMARY OF THE INVENTION

The method of the invention provides a process for the identification of interacting proteins that is suitable for high throughput analysis and amenable to automation. The present invention also allows for other properties of the proteins, and their interaction(s), to be characterized, including among other things, physical, structural and chemical properties, sequence information, and biological activity for the proteins alone and in complexes.

In part, the certain embodiments of the present invention use micro-columns to provide for high throughput methods. In another aspect, mass spectroscopy, in all its variations, may be used, which again may assist in achieving high throughput. In another aspect, the use of multiple ligand concentrations may be used to provide binding curves for certain embodiments. Using such multiple ligand concentrations may also allow for the reliability of the interactions that are identified to be confirmed.

The present invention achieve a number of desirable results and features, one or more of which (if any) may be present in any particular embodiment of the present invention: (i) interactions between two or more proteins may be identified by a variety of analytical means, including mass spectroscopy; (ii) certain methods of the present invention may be operated at high throughput; (iii) multiple concentration levels of protein ligands may be used to achieve more accurate results and provide additional information concerning the protein of interest and their interactions; and (iv) a variety of information may be obtained, including among other things, physical, structural and chemical properties, sequence information, and biological activity for the proteins alone and in complexes.

In certain embodiments, the identification of protein interactions is performed using affinity chromatography followed by mass spectrometric analysis.

In one such example, cellular extract or extracellular fluid may be loaded or otherwise added onto multiple experimental micro-columns or other appropriate vessels or wells, which have one or more bound ligand proteins. A control column, vessel or well without any bound ligand protein(s) may also be used. In certain examples, each of the experimental micro-columns, vessels or wells contains a different concentration of protein ligand bound to the matrix support. In certain embodiments, a fixed volume of cellular extract is chromatographed through each micro-column. In another aspect, affinity chromatography buffer (ACB) is chromatographed on a second control micro-column

which contains the highest concentration of ligand bound (coupled) to the matrix support. The components of the eluate may be separated, for example, on the basis of apparent molecular weight using SDS-PAGE, and visualized, for example, by protein staining. In this example, the interacting protein(s) (if any) may be observed to vary in amount in direct relation to the concentration of coupled protein ligand. The bands of interest may be excised from the gel and analyzed using mass spectrometric techniques.

In another aspect of the present invention, analytical techniques other than mass spectroscopy may be used to identify and otherwise characterize the components of the elute obtained from a method of the present invention.

In another aspect of the invention, kits containing some or all of the components necessary to complete a method of the present invention are provided.

In another aspect of the invention, apparatus necessary to conduct any of the methods are provided, including apparatus that may be used in a high throughput manner.

Generally, the nomenclature used herein and the laboratory procedures in spectroscopy, assays, drug discovery, cell culture, molecular genetics, protein purification, diagnostics, amino acid and nucleic acid chemistry described below are those well known and commonly employed in the art. The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, chemical syntheses, chemical analyses, biological assays, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, *Molecular Cloning A Laboratory Manual*, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press:1989); *DNA Cloning*, Volumes I and II (D. N. Glover ed., 1985); *Oligonucleotide Synthesis* (M. J. Gait ed., 1984); Mullis et al. U.S. Patent NO: 4,683,195; *Nucleic Acid Hybridization* (B. D. Hames & S. J. Higgins eds. 1984); *Transcription And Translation* (B. D. Hames & S. J. Higgins eds. 1984); *Culture Of Animal Cells* (R. I. Freshney, Alan R. Liss, Inc., 1987); *Immobilized Cells And Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide To Molecular Cloning* (1984); the treatise, *Methods In Enzymology* (Academic Press, Inc., N.Y.); *Gene Transfer Vectors For Mammalian Cells* (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); *Methods In Enzymology*, Vols. 154 and 155 (Wu et al. eds.), *Immunochemical Methods In Cell And Molecular Biology* (Mayer and Walker, eds., Academic Press, London, 1987); *Handbook Of Experimental Immunology*, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986);

Protein Purification: Principles and Practice, (R. K. Scopes, Third Edition, Springer Advanced Texts in Chemistry, 1994).

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a SDS-polyacrylamide gel run with the salt and SDS eluates from the affinity column using the *S. aureus* protein SA0005 as the ligand. The interacting protein is easily discerned from the background non-specific binding proteins as the band intensity increases with the increasing ligand concentration, but does not occur in the no-ligand and ACB controls.

Figure 2 is the mass spectrum of the tryptic peptides of the interacting protein excised from the gel of Figure 1. The technique used to obtain the spectrum is MALDI-TOF mass spectrometry.

The peptide masses were used to identify the interacting protein as a truncated form of EF-Tu.

Figure 3 is a SDS-polyacrylamide gel run with the salt and SDS eluates from the affinity column using the *S. aureus* protein SA0146 as the ligand. The interacting protein is easily discerned from the background non-specific binding proteins as the band intensity increases with the increasing ligand concentration, but does not occur in the no-ligand and ABC controls.

Figure 4 is the mass spectrum of the tryptic peptides of the interacting protein excised from the gel of Figure 3. The technique used to obtain the spectrum is MALDI-TOF mass spectrometry.

The peptide masses were used to identify the interacting protein as a conserved hypothetical protein of unknown function.

Figure 5 is a SDS-polyacrylamide gel run with the salt and SDS eluates from the affinity column using the *S. aureus* protein SA0203 as the ligand. The interacting protein is easily discerned from the background non-specific binding proteins as the band intensity increases with the increasing ligand concentration, but does not occur in the no-ligand and ACB controls.

Figure 6 is the mass spectrum of the tryptic peptides of the interacting protein excised from the gel of Figure 5. The technique used to obtain the spectrum is MALDI-TOF mass spectrometry.

The peptide masses were used to identify the interacting protein as a homologue of peptide chain release factor 3.

Figure 7 is a SDS-polyacrylamide gel run with the salt and SDS eluates from the affinity column using the *S. aureus* protein SA0276 as the ligand. The interacting proteins are easily discerned from the background non-specific binding proteins as the band intensities increases with the increasing ligand concentration, but do not occur in the no-ligand and ACB controls.

Figure 8 is the mass spectra of the tryptic peptides of the interacting proteins, interactor 1 and interactor 2, excised from the gel of Figure 7. The technique used to obtain the spectrum is MALDI-TOF mass spectrometry. The peptide masses from the respective spectra were used to identify the interacting proteins as homologues of glutamyl-tRNA Gln amidotransferase subunits A and B.

Figure 9 is a SDS-polyacrylamide gel run with the salt and SDS eluates from the affinity column using the *S. aureus* protein SA0526 as the ligand. The interacting protein is easily discerned from the background non-specific binding protein as the band intensity increases with the increasing ligand concentration, but does not occur in the no-ligand and ACB control.

Figures 10 is the mass spectra of the tryptic peptides of the interacting protein excised from the gel of Figure 9. The technique used to obtain the spectrum is MALDI-TOF mass spectrometry. The peptide masses were used to identify the interacting proteins as a homologue of EF-Tu.

Figure 11 is a polyacrylamide gel run with SDS eluates from the affinity column using the *S. aureus* protein SA0808 as the ligand. The interacting proteins are easily discerned from the background non-specific binding protein as the band intensity increases with the increasing ligand concentration, but does not occur in the no-ligand and ACB controls.

Figures 12a and 12b are the mass spectra of the tryptic peptides of the interacting proteins, interactor 1, interactor 2 (Figue12a), interactor 3 and interactor 4 (figure 12b), excised from the gel of Figure 11. The technique used to obtain the spectra is MALDI-TOF mass spectrometry. The peptide masses from the respective spectra were used to identify the interacting proteins as homologues of elongation factor G, trigger factor (prolyl isomerase), formate-tetrahydrofolate ligase, and EF-Tu.

Figure 13 is a polyacrylamide gel run with SDS eluates from the affinity column using the *S. aureus* protein SA0989 as the ligand. The interacting proteins are easily discerned from the background non-specific binding protein as the band intensity increases

with the increasing ligand concentration, but does not occur in the no-ligand and ACB controls.

Figure 14 is the mass spectra of the tryptic peptides of the interacting proteins, interactor 1 and interactor 3, excised from the gel of Figure 13. The technique used to obtain the spectra is MALDI-TOF mass spectrometry. The peptide masses from the respective spectra were used to identify two of the interacting proteins as homologues of trigger factor (prolyl isomerase) and enolase. The third is unidentified.

Figure 15 is a polyacrylamide gel run with SDS eluates from the affinity column using the unknown *S. aureus* protein SA1094 as the ligand. The interacting protein is easily discerned from the background non-specific binding protein as the band intensity increases with the increasing ligand concentration, but does not occur in the no ligand ACB controls.

Figure 16 is the mass spectrum of the tryptic peptides of the interacting protein excised from the gel of Figure 15. The technique used to obtain the spectrum is MALDI-TOF mass spectrometry.
The peptide masses were used to identify the interacting protein as a homologue of a putative peptidase.

Figure 17 is a polyacrylamide gel run with SDS eluates from the affinity column using the *S. aureus* protein SA1185 as the ligand. The interacting proteins are easily discerned from the background non-specific binding protein as the band intensity increases with the increasing ligand concentration, but does not occur in the no-ligand and ACB controls.

Figure 18 is the mass spectra of the tryptic peptides of the interacting proteins, interactor 1 and interactor 2, excised from the gel of Figure 17. The technique used to obtain the spectra is MALDI-TOF mass spectrometry. The peptide masses from the respective spectra were used to identify the interacting proteins as homologues of glucose-6-phosphate isomerase and cysteine synthetase.

Figure 19 is a polyacrylamide gel run with SDS eluates from the affinity column using the *S. aureus* protein SA1203 as the ligand. The interacting protein is easily discerned from the background non-specific binding protein as the band intensity increases with the increasing ligand concentration, but does not occur in the no-ligand and ACB controls.

Figure 20 is the mass spectrum of the tryptic peptides of the interacting protein excised from the gel of Figure 19. The technique used to obtain the spectrum is MALDI-TOF mass spectrometry.

The peptide masses were used to identify the interacting protein as a homologue of NADH dehydrogenase.

DETAILED DESCRIPTION

General Introduction

In part, the method of the invention uses a form of protein-affinity chromatography for the detection of protein-protein interactions and other protein information. In certain aspects, the methods of the invention allows for the isolation of specific protein interactors.

In certain embodiments, the interacting proteins are identified by protease digestion followed by mass spectrometry. During the past decade, new techniques in mass spectrometry have made it possible to accurately measure with high sensitivity the molecular weight of peptides and intact proteins. These techniques have made it much easier to obtain accurate peptide masses of a protein for use in databases searches. Mass spectrometry provides a method of protein identification that is both very sensitive (10 fmol - 1 pmol) and very rapid when used in conjunction with sequence databases. Advances in protein and DNA sequencing technology are resulting in an exponential increase in the number of protein sequences available in databases. As the size of DNA and protein sequence databases grows, protein identification by correlative peptide mass matching has become an increasingly powerful method to identify and characterize proteins.

Definitions

For convenience, certain terms employed in the specification, examples, and appended claims are collected here. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs.

The term "analyzing a protein by mass spectrometry" refers to using mass spectrometry to generate information which may be used to identify or aid in identifying a protein. Such information includes, for example, the mass or molecular weight of a protein, the amino acid sequence of a protein or protein fragment, a peptide map of a protein, and the purity or quantity of a protein.

An "agonist" increases, up regulates, mimics or potentiates by any means the biological activity of a polypeptide, nucleic acid, macromolecule, complex, molecule, species or the like.

The term "amino acid" is intended to embrace all molecules, whether natural or synthetic, which include both an amino functionality and an acid functionality and capable of being included in a polymer of naturally occurring amino acids. Exemplary amino acids include naturally occurring amino acids; analogs, derivatives and congeners thereof; amino acid analogs having variant side chains; and all stereoisomers of any of any of the foregoing.

The term "animal" refers to mammals, including, for example, humans, primates, bovines, ovines, porcines, canines, felines, and rodents (such as mice and rats).

An "antagonist" decreases, suppresses, down regulates or inhibits by any means the biological activity of a polypeptide, nucleic acid, macromolecule, complex, molecule, species or the like.

The term "binding" refers to an association, which may be a stable association, between two molecules, e.g., between a protein ligand and a another polypeptide, due to, for example, electrostatic, hydrophobic, ionic and/or hydrogen-bond interactions under physiological conditions.

The terms "biological activity" or "bioactivity" or "activity" or "biological function" refer to an effector or antigenic function that is directly or indirectly performed by a polypeptide, nucleic acid, macromolecule, complex, species or the like (whether in its native, denatured or other conformation).

"Cells," "host cells" or "recombinant host cells" are terms used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

The term "complex" refers to an association between at least two moieties (e.g. chemical or biochemical) that have an affinity for one another. Examples of complexes include associations between antigen/antibodies, lectin/avidin, target polynucleotide/probe oligonucleotide, antibody/anti-antibody, receptor/ligand, enzyme/ligand and the like. "Member of a complex" refers to one moiety of the complex, such as an antigen or ligand.

"Protein complex" or "polypeptide complex" refers to a complex comprising at least one polypeptide.

A "compound with therapeutic activity" refers to a therapeutic compound that binds to a polypeptide or other biological molecule, which may be naturally occurring, to alter or modulate its function.

The term "conserved residue" refers to an amino acid that is a member of a group of amino acids having certain common properties. The term "conservative amino acid substitution" refers to the substitution (conceptually or otherwise) of an amino acid from one such group with a different amino acid from the same group. A functional way to define common properties between individual amino acids is to analyze the normalized frequencies of amino acid changes between corresponding proteins of homologous organisms (Schulz, G. E. and R. H. Schirmer., Principles of Protein Structure, Springer-Verlag). According to such analyses, groups of amino acids may be defined where amino acids within a group exchange preferentially with each other, and therefore resemble each other most in their impact on the overall protein structure (Schulz, G. E. and R. H. Schirmer, Principles of Protein Structure, Springer-Verlag). One example of a set of amino acid groups defined in this manner include: (i) a charged group, consisting of Glu and Asp, Lys, Arg and His, (ii) a positively-charged group, consisting of Lys, Arg and His, (iii) a negatively-charged group, consisting of Glu and Asp, (iv) an aromatic group, consisting of Phe, Tyr and Trp, (v) a nitrogen ring group, consisting of His and Trp, (vi) a large aliphatic nonpolar group, consisting of Val, Leu and Ile, (vii) a slightly-polar group, consisting of Met and Cys, (viii) a small-residue group, consisting of Ser, Thr, Asp, Asn, Gly, Ala, Glu, Gln and Pro, (ix) an aliphatic group consisting of Val, Leu, Ile, Met and Cys, and (x) a small hydroxyl group consisting of Ser and Thr.

The term "DNA sequence encoding a polypeptide" may refer to one or more genes within an organism. As is well known in the art, genes for a particular polypeptide may exist in single or multiple copies within the genome of an organism. Such duplicate genes may be identical or may have certain modifications, including nucleotide substitutions, additions or deletions, which all still code for polypeptides having substantially the same activity. Moreover, certain differences in nucleotide sequences may exist between individual organisms, which are called alleles. Such allelic differences may result in differences in amino acid sequence of the encoded polypeptide yet still encode a protein with the same or substantially similar biological activity.

The term "domain" when used in connection with a polypeptide refers to a specific region within such polypeptide that comprises a particular structure or mediates a particular function.

A "fusion protein" or "fusion polypeptide" refers to a polypeptide comprising a first amino acid sequence encoding a polypeptide linked to at least one other amino acid sequence encoding another polypeptide that is not substantially homologous with any domain of the first polypeptide. The two polypeptide sequences may be linked in frame. A fusion protein may include a domain which is found (albeit in a different protein) in an organism which also expresses the first protein, or it may be an "interspecies", "intergenic", etc. fusion expressed by different kinds of organisms. In various embodiments, the fusion polypeptide may comprise one or more amino acid sequences linked to the first polypeptide. In the case where more than one amino acid sequence is fused to the first polypeptide, the fusion sequences may be multiple copies of the same sequence, or alternatively, may be different amino acid sequences. The fusion polypeptides may be fused to the N-terminus, the C-terminus, or the N- and C-terminus of the first polypeptide. Exemplary fusion proteins include polypeptides comprising a glutathione S-transferase tag (GST-tag), histidine tag (His-tag), maltose binding protein, an epitope for an available monoclonal antibody, an immunoglobulin domain or an immunoglobulin binding domain.

The term "gene" refers to a nucleic acid comprising an open reading frame encoding a polypeptide having exon sequences and optionally intron sequences. The term "intron" refers to a DNA sequence present in a given gene which is not translated into protein and is generally found between exons.

The term "having substantially similar biological activity", and like terms, refers to a biological activity of a first polypeptide which is substantially similar to at least one of the biological activities of a second polypeptide. A substantially similar biological activity means that the polypeptides carry out a similar function in the cell, e.g., a similar enzymatic reaction or a similar physiological process, etc. For example, two homologous proteins may have a substantially similar biological activity if they are involved in a similar enzymatic reaction, e.g., they are both kinases which catalyze phosphorylation of a substrate polypeptide, however, they may phosphorylate different regions on the same protein substrate or different substrate proteins altogether. Alternatively, two homologous proteins may also have a substantially similar biological activity if they are both involved in a similar physiological process, e.g., transcription. For example, two proteins may be

transcription factors, however, they may bind to different DNA sequences or bind to different polypeptide interactors. Substantially similar biological activities may also be associated with proteins carrying out a similar structural role in the cell, for example, two membrane proteins.

The term "isolated polypeptide" refers to a polypeptide, in certain embodiments prepared from recombinant DNA or RNA, or of synthetic origin, or some combination thereof, which (1) is not associated with proteins that it is normally found with in nature, (2) is isolated from the cell in which it normally occurs, (3) is isolated free of other proteins from the same cellular source, (4) is expressed by a cell from a different species, or (5) does not occur in nature.

The term "isolated nucleic acid" refers to a polynucleotide of genomic, cDNA, or synthetic origin or some combination there of, which (1) is not associated with the cell in which the "isolated nucleic acid" is found in nature, or (2) is operably linked to a polynucleotide which it is not linked to in nature.

The terms "label" or "labeled" refer to incorporation of a detectable marker into a molecule, such as a polypeptide. Various methods of labeling polypeptides are known in the art and may be used. Examples of labels for polypeptides include, but are not limited to, the following: radioisotopes, fluorescent labels, heavy atoms, enzymatic labels or reporter genes, chemiluminescent groups, biotinyl groups, predetermined polypeptide epitopes recognized by a secondary reporter (e.g., leucine zipper pair sequences, binding sites for secondary antibodies, metal binding domains, epitope tags). Examples and use of such labels are described in more detail below. In some embodiments, labels are attached by spacer arms of various lengths to reduce potential steric hindrance.

The terms "ligand", "protein ligand" or "bait" refer to a polypeptide or other biological material which is used as a target to find other proteins which may associate with it. In certain embodiments, a bait protein is tagged or immobilized. The use of protein ligands in the present invention is described in more detail below.

The term "modulation", when used in reference to a functional property or biological activity or process (e.g., enzyme activity or receptor binding), refers to the capacity to either up regulate (e.g., activate or stimulate) or down regulate (e.g., inhibit or suppress) such property, activity or process. In certain instances, such regulation may be contingent on the occurrence of a specific event, such as activation of a signal transduction pathway, and/or may be manifest only in particular cell types.

The term "modulator" refers to a polypeptide, nucleic acid, macromolecule, complex, molecule, small molecule, species or the like (naturally occurring or non-naturally occurring), or an extract made from biological materials such as bacteria, plants, fungi, or animal cells or tissues, that may be capable of causing modulation. The activity of a modulator may be known, unknown or partially known. In certain instances, a modulator may interfere with the binding between a polypeptide or other biological material and a protein ligand.

The term "motif" refers to an amino acid sequence that is commonly found in a protein of a particular structure or function. Typically a consensus sequence is defined to represent a particular motif. The consensus sequence need not be strictly defined and may contain positions of variability, degeneracy, variability of length, etc. The consensus sequence may be used to search a database to identify other proteins that may have a similar structure or function due to the presence of the motif in its amino acid sequence. For example, on-line databases may be searched with a consensus sequence in order to identify other proteins containing a particular motif. Various search algorithms and/or programs may be used, including FASTA, BLAST or ENTREZ. FASTA and BLAST are available as a part of the GCG sequence analysis package (University of Wisconsin, Madison, Wis.). ENTREZ is available through the National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD.

The term "naturally-occurring", as applied to an object, refers to the fact that an object may be found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism (including bacteria) that may be isolated from a source in nature and which has not been intentionally modified by man in the laboratory is naturally-occurring.

The term "nucleic acid", which is used herein interchangeably with "polynucleotides", refers to a polymeric form of nucleotides, either ribonucleotides or deoxynucleotides or a modified form of either type of nucleotide. The terms should also be understood to include, as equivalents, analogs of either RNA or DNA made from nucleotide analogs, and, as applicable to the embodiment being described, single-stranded (such as sense or antisense) and double-stranded polynucleotides.

The term "operably linked", when describing the relationship between two nucleic acid regions, refers to a juxtaposition wherein the regions are in a relationship permitting them to function in their intended manner. For example, a control sequence "operably

linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences, such as when the appropriate molecules (e.g., inducers and polymerases) are bound to the control or regulatory sequence(s).

The terms "pharmaceutical agent" or "drug" refer to a compound or composition capable of inducing a desired therapeutic effect when properly administered to a patient.

The term "phenotype" refers to the entire physical, biochemical, and physiological makeup of a cell, e.g., having any one trait or any group of traits.

The term "polypeptide", and the terms "protein" and "peptide" which are used interchangeably herein, refers to a polymer of amino acids. Exemplary polypeptides include gene products, naturally occurring proteins, homologs, orthologs, paralogs, fragments, and other equivalents and analogs of the foregoing.

The term "polypeptide fragment", when used in reference to a reference polypeptide, refers to a polypeptide in which amino acid residues are deleted as compared to the reference polypeptide itself, but where the remaining amino acid sequence is usually identical to the corresponding positions in the reference polypeptide. Such deletions may occur at the amino-terminus or carboxy-terminus of the reference polypeptide. Fragments typically are at least 5, 6, 8 or 10 amino acids long, at least 14 amino acids long, at least 20, 30, 40 or 50 amino acids long, at least 75 amino acids long, or at least 100, 150, 200, 300, 500 or more amino acids long.

The term "purified" refers to an object species that is the predominant species present (i.e., on a molar basis it is more abundant than any other individual species in the composition). A "purified fraction" is a composition wherein the object species comprises at least about 50 percent (on a molar basis) of all species present. In making the determination of the purity of a species in solution or dispersion, the solvent or matrix in which the species is dissolved or dispersed is usually not included in such determination; instead, only the species (including the one of interest) dissolved or dispersed are taken into account. Generally, a purified composition will have one species that comprises more than about 80 percent of all species present in the composition, more than about 85%, 90%, 95%, 99% or more of all species present. The object species may be purified to essential homogeneity (contaminant species cannot be detected in the composition by conventional detection methods) wherein the composition consists essentially of a single species. A skilled artisan may purify a polypeptide using standard techniques for protein purification

in light of the teachings herein. Purity of a polypeptide may be determined by a number of methods known to those of skill in the art, including for example, amino-terminal amino acid sequence analysis, gel electrophoresis, mass-spectrometry analysis and the methods described in the Exemplification section herein.

The terms "recombinant protein" or "recombinant polypeptide" refer to a polypeptide which is produced by recombinant DNA techniques. An example of such techniques includes the case when DNA encoding the expressed protein is inserted into a suitable expression vector which is in turn used to transform a host cell to produce the protein or polypeptide encoded by the DNA.

The term "regulatory sequence" is a generic term used throughout the specification to refer to polynucleotide sequences, such as initiation signals, enhancers, and promoters, that are necessary or desirable to effect the expression of coding and non-coding sequences to which they are operably linked.

The term "reporter gene" refers to a nucleic acid comprising a nucleotide sequence encoding a protein that is readily detectable either by its presence or activity, including, but not limited to, luciferase, fluorescent protein (e.g., green fluorescent protein), chloramphenicol acetyl transferase, ss-galactosidase, secreted placental alkaline phosphatase, ss-lactamase, human growth hormone, and other secreted enzyme reporters.

The term "sequence homology" refers to the proportion of base matches between two nucleic acid sequences or the proportion of amino acid matches between two amino acid sequences. When sequence homology is expressed as a percentage, e.g., 50%, the percentage denotes the proportion of matches over the length of sequence from a desired sequence that is compared to some other sequence. Gaps (in either of the two sequences) are permitted to maximize matching; gap lengths of 15 bases or less are usually used, 6 bases or less are used more frequently, with 2 bases or less used even more frequently. The term "sequence identity" means that sequences are identical (i.e., on a nucleotide-by-nucleotide basis for nucleic acids or amino acid-by-amino acid basis for polypeptides) over a window of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical amino acids occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying

the result by 100 to yield the percentage of sequence identity. Methods to calculate sequence identity are known to those of skill in the art.

The term "small molecule" refers to a compound, which has a molecular weight of less than about 5 kD and most preferably less than about 2.5 kD. Small molecules may be nucleic acids, peptides, polypeptides, peptidomimetics, carbohydrates, lipids or other organic (carbon containing) or inorganic molecules. Many pharmaceutical companies have extensive libraries of chemical and/or biological mixtures, often fungal, bacterial, or algal extracts, which may be used in an assay of the present invention. The term "small organic molecule" refers to a small molecule that is often identified as being an organic or medicinal compound, and does not include molecules that are exclusively nucleic acids, peptides or polypeptides.

The terms "solid support", "matrix," "matrix support," used interchangeably, refers to a material which is an insoluble matrix, and may (optionally) have a rigid or semi-rigid surface. Such materials may take the form of small beads, pellets, disks, chips, dishes, multi-well plates, wafers or the like, although other forms may be used. The term "column support" is an example of a solid support, in which the insoluble matrix is arranged in a column or other shape that facilitates the performance of the inventive methods. In some embodiments, at least one surface of the substrate will be substantially flat. The term "surface" refers to any generally two-dimensional structure on a solid substrate and may have steps, ridges, kinks, terraces, and the like without ceasing to be a surface.

The term "soluble support" refers to a material that is at least partially soluble in some or all of the conditions in which it will be used. A support is termed a "soluble support" if the support, or the support with a protein ligand or other chemical moiety(ies) immobilized thereto, is soluble under one or more of the conditions employed. In certain instances, a soluble support may be rendered insoluble under defined conditions. Accordingly, a soluble support may be soluble under certain conditions and insoluble under other conditions. Examples of soluble supports include certain polymers, such as polyethylene glycols or polyvinyl alcohols.

The terms "immobilized" or "coupling," used with respect to a species, refer to a condition in which the species is attached to a surface with an attractive force stronger than attractive forces that are present in the intended environment of use of the surface, and that act on the species. As one example of such immobilization or coupling, a protein ligand

may be immobilized or coupled on a solid support by one of the methods described in detail below.

The term "soluble" as used herein with reference to a polypeptide, means that upon expression in cell culture, at least some portion of the polypeptide expressed remains in the cytoplasmic fraction of the cell and does not fractionate with the cellular debris upon lysis and centrifugation of the lysate. Solubility of a polypeptide may be increased by a variety of art recognized methods, including fusion to a heterologous amino acid sequence, deletion of amino acid residues, amino acid substitution (e.g., enriching the sequence with amino acid residues having hydrophilic side chains), and chemical modification (e.g., addition of hydrophilic groups). The solubility of polypeptides may be measured using a variety of art recognized techniques, including, dynamic light scattering to determine aggregation state, UV absorption, centrifugation to separate aggregated from non-aggregated material, and SDS gel electrophoresis (e.g., the amount of protein in the soluble fraction is compared to the amount of protein in the soluble and insoluble fractions combined). Polypeptides may be at least about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% or more soluble, e.g., at least about 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% or more of the total amount of protein expressed in the cell is found in the cytoplasmic fraction.

The term "specifically hybridizes" refers to detectable and specific nucleic acid binding. Polynucleotides, oligonucleotides and nucleic acids selectively hybridize to nucleic acid strands under hybridization and wash conditions that minimize appreciable amounts of detectable binding to nonspecific nucleic acids. High stringency conditions may be used to achieve selective hybridization conditions as known in the art and discussed herein. Generally, the nucleic acid sequence homology between polynucleotides, oligonucleotides, and nucleic acids will be at least 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 98%, 99%, or more. In certain instances, hybridization and washing conditions are performed at high stringency according to conventional hybridization procedures.

As applied to proteins, the term "substantial identity" means that two protein sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap weights, typically share at least about 70 percent sequence identity, alternatively at least about 80, 85, 90, 95 percent sequence identity or more. In certain instances, residue positions that are not identical differ by conservative amino acid substitutions.

The term "test compound" refers to a molecule to be tested by one or more screening method(s) as a putative modulator of a polypeptide or other biological material. A test compound is usually not known to bind to a target of interest. The term "novel test compound" refers to a test compound that is not in existence as of the filing date of this application.

The term "treating" is intended to encompass curing as well as ameliorating at least one symptom of a condition or disease.

The term "vector" refers to a nucleic acid capable of transporting another nucleic acid to which it has been linked. Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of "plasmids" which refer to circular double stranded DNA molecules which, in their vector form are not bound to the chromosome.

Protein ligand

In one aspect, a protein ligand is immobilized on the solid support or other support and used as a target to find other proteins or other biological materials which may associate with it. The possible protein ligands include, among others, naturally occurring proteins, modified proteins, synthetic proteins and subdomains or fragments of proteins.

In certain embodiments, the protein to be used as the ligand should be purified. In certain embodiments, the ligand is at least 90% purified. Such high purity makes it more likely that the interacting proteins that are detected are binding to the intended ligand rather than a contaminant.

In one aspect, a method of obtaining protein, if the gene is available, is through the use of fusion proteins. If, for technical reasons, an impure ligand must be used, it may be helpful to use a control preparation that mimics the contaminants but does not contain ligand. In an illustrative embodiment, a fusion protein may be provided which adds a domain that permits the protein to be bound to an insoluble matrix.

The ligand protein to be used for affinity chromatography may be encoded by the nucleic acid of a virus or any other organism. The nucleic acid fragment to be cloned may be identified from the gene sequence when the genome of the organism is partly or entirely known. Isolation of the nucleic acid fragment is performed, for example, by gel electrophoresis after digestion with a restriction enzyme, by random fragmentation or by amplification from genomic DNA or other nucleic acid or a recombinant clone by using the

polymerase chain reaction (PCR). Other methods for obtaining suitable nucleic acid for a protein ligand are known in the art and incorporated by reference below.

DNA encoding the protein or protein fragment may be cloned into an expression vector. The wide availability of recombinant technology makes it feasible to generate expression systems that may be able to produce sufficient quantities of a selected protein for use as a ligand in the method of the invention.

As an illustrative example, the steps for protein production include: generation of the protein expression systems, over-expressing the protein and purifying the protein. The generation of a clone for any particular gene of interest, and its incorporation into a suitable expression vector, is now a straightforward task. In certain examples, it may be done in a parallel fashion for high throughput production. Edwards et al., Nature Structural Biology 2000, 7, 970-972.

The selection of target proteins from partially or completely sequenced genomes may take advantage of the availability of these cloned genes. However, even if a clone of a particular protein of interest is not readily available, those of skill in the art may be able to generate a cDNA clone or other nucleic acid clone without undue experimentation.

In certain embodiments, to obtain expression of a cloned nucleic acid, the expression vector for expression in bacteria typically comprises a strong promoter to direct transcription, a transcription/translation terminator, and if the nucleic acid encodes a peptide or polypeptide, a ribosome binding site for translational initiation. Suitable bacterial promoters are well known in the art and described, e.g., in Sambrook et al. and Ausubel et al. Bacterial expression systems are available in, e.g., E. coli, Bacillus sp., and Salmonella (Palva et al., Gene 22:229-235 (1983); Mosbach et al., Nature 302:543-545 (1983). Kits for such expression systems are commercially available.

Post-translational modification of the ligand protein may be related to the protein's ability to interact with other proteins. In certain cases, eukaryotic expression systems may be preferred, where post-translational modifications are important, for example, glycosylation. Eukaryotic expression systems for mammalian cells, yeast, and insect cells are well known in the art and are also commercially available. In some cases, it may be preferable to employ expression vectors which may be propagated in both prokaryotic and eukaryotic cells, enabling, for example, nucleic acid purification and analysis using one organism and protein expression using another.

Transfection methods used to produce bacterial, mammalian, yeast or insect cells or cell lines that express large quantities of protein are well known in the art. These include the use of calcium phosphate transfection, polybrene, protoplast fusion, electroporation, liposomes, microinjection, viral vectors and any of the other well known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (see, e.g., Sambrook et al., supra). In some of those examples, after the expression vector is introduced into the cells, the transfected cells may be cultured under conditions favoring expression of protein, which may then be purified using standard techniques.

The protein may be expressed in suitable amounts for use as the ligand. There are several expression systems that have been extensively studied, and others will be developed and be of use in the present invention. Some of these include: 1) bacterial (*E. coli*), 2) methylotrophic yeast (*Pichia pastorisis*), 3) viral (baculovirus, adenovirus, vaccinia and some RNA viruses), 4) cell culture (mammalian and insect), and 5) *in vitro* translation. Although the expression of any particular protein may be idiosyncratic, the availability of these and other expression systems significantly increases the ability to produce quantities of protein adequate to perform the present invention.

In situations in which relatively large amounts of relatively pure protein in native form are required, it may be desirable to employ expression systems characterized by high expression levels and efficient protein processing, including cleavage of signal peptides and other post-translational modifications. For example, the baculovirus expression system is widely used to express a variety of proteins in large quantities. In addition to fulfilling the above requirements, the size of the expressed protein is not limited, and expressed proteins are typically correctly folded and in a biologically active state. Baculovirus expression vectors and expression systems are commercially available (Clontech, Palo Alto, CA; Invitrogen Corp., Carlsbad, CA).

In another aspect, once a protein has been expressed to an acceptable level, the protein may be purified from the other contents of the cell system that was utilized for expression. The proteins may be expressed fused to tags that aid subsequent purification or measurement techniques. Typical tags bind specifically to particular affinity matrices, allowing the attached protein to be purified without regard to its physical or biochemical characteristics. Such tags may then be cleaved, leaving the protein in its native form. Examples of tags include histidine rich sequences which bind to various metal ions,

glutathione-S-transferase (GST) tags which selectively bind to glutathione, maltose-binding protein, or an epitope for an available monoclonal antibody, and other suitable tags are known to those of skill in the art.

In certain embodiments, the recombinant protein to be used as a ligand may be purified from the cells of the heterologous system by a chromatographic procedure that makes use of the tag on the protein. Examples of such procedures include, but are not limited to, nickel chelate chromatography, chromatography on a glutathione column, or chromatography on a suitable antibody column. In certain cases, the fusion protein also includes a cleavable sequence of amino acids between the protein of interest and the tag sequence whereby the tag can be cleaved from the protein of interest. Typically, this is accomplished with a protease that cleaves the sequence under conditions where the protein of interest is not degraded, or with an intein sequence, which allows for internal cleavage of the protein. Alternatively, the tags provide a method for specifically anchoring proteins to a solid support. In another alternative, the protein ligands may contain the expression and/or purification tags.

In still another aspect, the ligand protein may be purified by other acceptable methods known in the art, for example by immuno-chromatographic methods. Specific antibodies that recognize the ligand protein may be generated in a number of organisms using ligand protein, or a portion of it. The antibodies may be linked to a solid support and used to purify the ligand protein from a cellular extract or other source.

In those methods of the invention using a solid support, a ligand protein may be attached by a variety of means known to those of skill in the art. For example, the ligand protein may be coupled directly (through a covalent linkage) to commercially available pre-activated resins as described in Formosa et al., Methods in Enzymology 1991, 208, 24-45; Sopta et al, J. Biol. Chem. 1985, 260, 10353-60; Archambault et al., Proc. Natl. Acad. Sci. USA 1997, 94, 14300-5. Alternatively, the ligand protein may be tethered to the solid support through high affinity binding interactions. If the ligand is expressed fused to a tag, such as GST, the fusion tag can be used to anchor the ligand protein to the matrix support, for example Sepharose beads containing immobilized glutathione. Solid supports that take advantage of these tags are commercially available.

In another aspect, the support to which a protein ligand may be immobilized is a soluble support, which may facilitate certain steps performed in the methods of the present invention. For example, the soluble support may be soluble in the conditions employed to

crated a binding interaction between a target and the protein ligand, and then used under conditions in which it is a solid for elution of the proteins or other biological materials that bind to protein ligand.

In certain embodiments, the ligand protein may be coupled to the matrix or other solid support by a covalent linkage. The coupling procedures may make use of the many primary amino groups (lysines and the amino-terminal residues) which may be on the surface of the protein. Any coupling chemistry which makes use of primary amines is appropriate. In addition, a different reactive chemical moiety may be used which reacts at a reasonable rate at the physiological pH (e.g., N-hydroxy-succinimide works well at pH 7.5-8.0). Commercially available solid supports have reactive moieties for coupling to proteins, for example, cyanogen bromide-activated Sepharose (Pharmacia) or N-hydroxysuccinimide-activated agarose matrix, available as Affi-Gel 10 (Bio-Rad).

In certain instances, failure to detect an interacting protein may result from inactivation of the ligand protein during coupling to the solid support. To minimize such an occurrence, one would usually like to have a ligand protein randomly tethered to the matrix through one covalent bond. When the ligand is attached randomly, it is believed that some of the immobilized protein molecules will always be oriented in such a way as to be able to interact with the proteins in the extract.

In certain embodiments, the ligand protein may be contacted with the matrix under conditions that are favorable for coupling. For example, if the matrix is a bead, the solid support beads are mixed and shaken gently, tumbled or rotated with solution containing the protein ligand. Alternatively, the protein ligand solution is reacted with the activated solid support which is already packed into a column. The latter method, using a pre-packed column, may have certain advantages over other methods, as it typically uses less ligand and is amenable to automation and high throughput analysis. The concentration of salt and the pH may need to be adjusted to be appropriate for the resin and the protein ligand that are being used.

To achieve optimal sensitivity for the inventive methods, it may be important to choose a matrix that will couple a maximum concentration of protein ligand without introducing potentially denaturing multiple cross-links to individual proteins, or otherwise materially interfered with the binding of polypeptide or other moieties to the immobilized protein ligand. One element to consider in choosing a matrix is minimizing the non-specific interactions between proteins from the extract and the matrix. The matrix support

may be chosen from, for example, agarose, sepharose, glass beads, latex beads, cellulose, or dextran.

The concentration of the coupled protein ligand may have an affect on the sensitivity of the inventive methods. For examples, we have observed that in certain embodiments, to detect interactions most efficiently, the concentration of the ligand protein bound to the matrix should be at least 10-fold higher than the K_d of the interaction. Thus, the concentration of the ligand protein bound to the matrix should be highest for the detection of the weakest protein-protein interactions. However, if the concentration of the immobilized protein ligand is not as high as may be ideal, it may still be possible to observe protein-protein interactions of interest by, for example, increasing the concentration of the polypeptide or other moiety that interacts with the coupled protein ligand. The level of detection will of course vary with each different protein ligand, interactor, conditions of the assay, etc.

In another aspect, the coupling may be done at various ratios of the protein ligand to the resin. An upper limit of the protein : resin ratio may be determined by the isoelectric point and the ionic nature of the protein, although it may be possible to achieve higher protein ligand concentrations by use of various methods.

In certain embodiments, several concentrations of the protein ligand immobilized on a solid or soluble support may be used. One advantage of using multiple concentrations, although not a requirement, is that one may be able to obtain an estimate for the strength of the protein-protein interaction that is observed in the affinity chromatography experiment, described in detail below. Another advantage of using multiple concentrations is that a binding curve which has the proper shape may indicate that the interaction that is observed is biologically important rather than a spurious interaction with denatured protein. For these two reasons, and others, a number of embodiments of the present invention as described in the Exemplification section below use solid supports with varying concentrations of immobilized protein ligand.

In one example of such an embodiment, a series of columns may be prepared with varying concentrations of protein ligand (mg protein ligand/ml resin volume). The number of columns employed may be between 2 to 8, 10, 12 or 15, each with a different concentration of attached ligand. Larger numbers of columns may be used if appropriate for the protein ligand being examined, and multiple columns may be used with the same concentration as any methods may require. In certain embodiments, 4 to 6 columns are

prepared with varying concentrations of ligand. In another aspect of this embodiment, two control columns may be prepared: one that contains no ligand and a second that contains the highest concentration of ligand but is not treated with extract. After elution of the columns and separation of the eluent components (by one of the methods described below), it may be possible to distinguish the interacting proteins (if any) from the non-specific bound proteins as follows. The concentration of the interacting proteins, as determined by the intensity of the band on the gel, will increase proportionally to the increase in protein ligand concentration but will be missing from the second control column. This allows for the identification of unknown interacting proteins.

The coupling of the protein with the solid support may be terminated, if desired, but not necessarily, by reacting the support with ethanolamine. It has been standard practice to treat the column support resin with ethanolamine and bovine serum albumin (BSA) after the ligand protein is coupled. This was done to block the remaining reactive groups on the resin. We have found that it is preferable to avoid the treatment of the resin with BSA and ethanolamine. By omitting this treatment, we have found that the non-specific binding of proteins from an extract to the resin is reduced by about five-fold.

Micro-columns

The method of the invention may be used for small-scale analysis. A variety of column sizes, types, and geometries may be used. In addition, other vessel shapes and sizes having a smaller scale than is usually found in laboratory experiments may be used as well, including a plurality of wells in a plate.

For high throughput analysis, it is advantageous to use small volumes, about 20, 30, 50, 80 or 100 µl. Larger or small volumes may be used, as necessary, and it may be possible to achieve high throughput analysis using them.

In one example, a column may be constructed in a glass capillary with a drawn-out tip or a plastic pipette tip. In order to retain the solid support in the capillary or pipette tip, the tip may blocked with glass beads, glass wool, filter paper, a frit or other material that blocks the solid support and is permeable to liquids. The entire affinity chromatography procedure may be automated by assembling the micro-columns into an array (e.g. with 96 micro-column arrays). By the term "array," it is understood to mean a collection of multiple micro-columns or other vessels. In certain arrays, all the micro-columns or other vessels are of the approximate same dimension and scale. For other arrays, the micro-columns and other vessels are physically disposed in a manner that allows all of them to be

used at approximately the same time (although not all such columns or vessels need to be used at the same time). The number of columns or vessels in an array is usually more than ten, and may number from to 20, 50, 100 etc. One example of a vessel is a well in a plate. A 96-well plate is one example of an array. Another example of an array is a "multi-well platform," which has a plurality of wells within a frame.

Preparation of Extracts and Other Materials for Analysis

Any type of suitable mixtures may be used to analyze for protein-protein interactions, provided it contains one or more proteins or other biological materials that may interact with the protein ligand. For purposes of this invention, the term "extract" encompasses all such mixtures, whether extracted from a biological source or not. For example, a suitable mixture that is an extract hereunder is a solution containing two polypeptides prepared to be used in this invention.

In one aspect, the extract may be a cellular extract or extracellular fluid. In one aspect, the extract contains a mixture of proteins derived from a natural source, as well as possibly other materials derived from a natural source. More generally, suitable extracts are made from biological materials such as bacteria, plants, fungi, or cells or tissues.

In general, the choice of starting material for the extract is based upon the cell or tissue type or type of fluid that would be expected to contain proteins that interact with the target protein. For example, micro-organisms or other organisms are grown in a medium that is appropriate for that organism and can be grown in specific conditions to promote the expression of proteins that may interact with the target protein.

Exemplary starting materials that may be used to make the extract include: 1) one or more types of tissue derived from an animal, plant, or other multi-cellular organism, 2) cells grown in tissue culture that were derived from an animal or human, plant or other source, 3) micro-organisms grown in suspension or non-suspension cultures, 4) virus-infected cells, 5) purified organelles (including, but not restricted to nuclei, mitochondria, membranes, Golgi, endoplasmic reticulum, lysosomes, or peroxisomes) prepared by differential centrifugation or another procedure from animal, plant or other kinds of eukaryotic cells, 6) serum or other bodily fluids including, but not limited to, blood, urine, semen, synovial fluid, cerebrospinal fluid, amniotic fluid, lymphatic fluid or interstitial fluid.

In certain embodiments, whole cell extracts may be used as the source of interacting proteins. Alternatively, in some cases, a total cell extract may not be the optimal source of

interacting proteins. For example, if the ligand is known or thought to act in the nucleus, a nuclear extract can provide a 10-fold enrichment of proteins that are likely to interact with the ligand. In addition, proteins that are present in the extract in low concentrations may be enriched using another chromatographic method to fractionate the extract before screening various pools for an interacting protein.

One way to use whole cell extracts follows. Any of the techniques described may be used alone or with other techniques to prepare suitable extracts for the inventive methods. The cells are lysed by standard methods, including, but not limited to enzymatic lysis, grinding with alumina or another abrasive, use of a French pressure cell, sonication, treatment with detergent, beating with glass beads in a bead beater or blender, cryogenic grinding, exposure to differential osmotic pressure, use of a mill, or use of a Dounce homogenizer. It may be advantageous to carry out the procedure at a low temperature (e.g., 4°C) in order to retard denaturation or degradation of proteins in the extract, although it may not be necessary. Next, tissue or cells or cell extract is suspended in a solution containing Tris or Hepes or another biological buffer that is standard in the art at a concentration that is adequate to establish the pH of the extract. The pH is adjusted to be appropriate for the body fluid or tissue, cellular, or organellar source that is used for the procedure (e.g. pH 7-8 for cytosolic extracts from mammals, but low pH for lysosomal extracts). Next, the concentration of chaotropic or non-chaotropic salts in the extracting solution may need to be adjusted so as to extract the appropriate sets of proteins for the procedure. Glycerol may be added to the lysate, as it aids in maintaining the stability of many proteins and also reduces background non-specific binding. Both the lysis buffer and column buffer may contain protease inhibitors to minimize proteolytic degradation of proteins in the extract and to protect the ligand. Appropriate co-factors that could potentially interact with the interacting proteins may be added to the extracting solution. One or more nucleases or another reagent is added to the extract, if appropriate, to prevent protein-protein interactions that are mediated by nucleic acids. Appropriate detergents or other agents are added to the solution, if desired, to extract membrane proteins from the cells or tissue. A reducing agent (e.g. dithiothreitol or 2-mercaptoethanol or glutathione or other agent) may be added to extracts derived from cells, but may be omitted when the source of protein extract is derived from an extracellular source. Trace metals or a chelating agent may be added, if desired, to the extracting solution.

Next, the extract is centrifuged in a centrifuge or ultracentrifuge or filtered to provide a clarified supernatant solution. This supernatant solution may be dialyzed using dialysis tubing, or another kind of device that is standard in the art, against a solution that is similar to, but may not be identical with, the solution that was used to make the extract. An example of a change in the dialysis solution is to adjust the concentrations of salts to the ones that will be used for the affinity chromatography procedure. The dialysis procedure may last from less than an hour to many hours and can be omitted for fluids derived from extracellular sources or, in some cases, for extracts derived from intracellular sources. After dialysis, the extract containing proteins may be used immediately, stored for a short time, stored for many hours at a low temperature or stored in a frozen state at a low temperature (e.g., -80°C). The extract may be clarified by centrifugation or filtration again immediately prior to its use in affinity chromatography.

In some cases, the crude lysate or other material used may contain small molecules that may interfere with the affinity chromatography. This may be remedied by precipitating proteins with ammonium sulfate, centrifugation of the precipitate, and re-suspending the proteins in the affinity column buffer followed by dialysis. An additional centrifugation of the sample may be needed to remove any particulate matter prior to application to the affinity columns.

The amount of extract applied to the column is important for two opposing reasons. If too little extract is applied to the column and the interacting protein is present at low concentration, the level of interacting protein retained by the column may be difficult to detect. Conversely, if too much extract is applied to the column, protein may precipitate on the column or competition by abundant interacting proteins for the limited amount of protein ligand may result in a difficulty in detecting minor species. The appropriate amount of extract may be adjusted as is appropriate for the extract, protein ligand, support and other parameters of any embodiment of the present invention.

Affinity Chromatography

This section describes in general a variety of methods for completing affinity chromatography as used in the present invention. After completing the affinity chromatography, elutions or eluates will be obtained that may contain one or more proteins or other biological materials that interact with the protein ligand, which proteins or other biological materials may be subjected to analysis as described herein.

In one example, the columns may be loaded with extract from an appropriate source, which may have been dialyzed against a buffer that is consistent with the nature of the expected interaction. Glycerol may be included in the buffer. Any standard biological buffer can be used. The pH, salt concentrations and the presence or absence of reducing and chelating agents, trace metals, detergents, and co-factors may be adjusted according to the nature of the expected interaction. Usually, the pH and the ionic strength are chosen so as to be close to physiological for the source of the extract. In certain examples, the extract is loaded under gravity onto the columns at a flow rate of about 4-6 column volumes per hour, but this flow rate may be adjusted for particular circumstances, such as for an automated procedure.

The volume of the extract that is loaded on the columns may be varied, but is most commonly equivalent to about 5 to 10 column volumes, but may be 1, 3, 15, or even 20 times the column volumes. When large volumes of extract are loaded on the columns, it has been observed that there is an improvement in the signal-to-noise ratio because more protein from the extract is available to bind to the protein ligand, whereas the background binding of proteins from the extract to the solid support saturates with low amounts of extract. Alternatively, the appropriate volume may depend on the support, protein ligand, the size, shape and other characteristics of the column, vessel or array used, and other features of the method being practiced.

In certain embodiments, a control column is included that contains at least the highest concentration of protein ligand, but buffer rather than extract is loaded onto this column. Usually, the elutions (eluates) from this column will contain ligand protein that failed to be attached to the column in a covalent manner, but no proteins that are derived from the extract.

In certain instances, after the extract is applied to the columns, the columns are washed with a buffer appropriate to the nature of the interaction being analyzed, usually, but not necessarily, the same as the loading buffer. An elution buffer with an appropriate pH, glycerol, and the presence or absence of reducing agent, chelating agent, cofactors, and detergents are all important considerations. The columns are washed with about 5 to 20 column volumes of each wash buffer to eliminate unbound proteins from the natural extract. The flow rate of the wash is usually adjusted to about 4 to 6 column volumes per hour by using gravity or an automated procedure, but other flow rates are possible in

specific circumstances. As described above, the volume and flow rate may be varied as required.

In order to elute the proteins that have been retained by the column, the interactions between the extract proteins and the column ligand are disrupted. There are a number of ways known to those in the art to achieve such a disruption. By way of example, this disruption may achieved by eluting the column with a solution of salt or detergent. In certain cases, retention of activity by the eluted proteins requires the presence of glycerol and a good buffer of appropriate pH, as well as proper choices of ionic strength and the presence or absence of appropriate reducing agent, chelating agent, trace metals, cofactors, detergents, chaotropic agents, and other reagents.

In another aspect, if physical identification of the bound proteins is the objective, the elution may be performed sequentially, first with buffer of high ionic strength and then with buffer containing a protein denaturant, most commonly, but not restricted to sodium dodecyl sulfate (SDS), urea, or guanidine hydrochloride. We have found that, in certain embodiments, it is advantageous to simply elute the column with a protein denaturant, particularly SDS, for example as a 1% SDS solution. Using only the SDS wash, and omitting the salt wash results in SDS-gels that have higher resolution (sharper bands with less smearing). This makes it easier to visualize specifically bound proteins against the background of non-specifically bound proteins. In addition, using only the SDS wash results in half as many samples to analyze by electrophoresis. The number of samples to be analyzed is an important consideration for the development of high throughput techniques.

The volume of the eluting solution may be varied but is normally about 2 to 4 column volumes. For example, for 20 ml columns, the flow rate of the eluting procedures are most commonly about 4 to 6 column volumes per hour, under gravity, but can be varied in an automated procedure. As before, the volume and flow rate may be adjusted as appropriate.

In another aspect, the present invention contemplates including modulators to affect any protein-protein interaction that would otherwise occur during affinity chromatography. By this method, a modulator is included in a mixture and the results of including such modulator are compared to the results when no modulator (or a different one) is included in the mixture. Any decrease in binding may indicate that the modulator interferes with the binding of a polypeptide or other biological material to the protein ligand. The modulators (e.g., antagonists and agonists) identified by such a method may be employed, for instance,

to treat a disease or condition of a patient (including humans and animals). In another embodiment, modulators identified by methods of the present invention may be used in the manufacture of a medicament for any number of uses, including, for example, treating any disease or other condition of a patient.

Separation of Eluent Components

There are number of methods in the art that may be used to separate any of the proteins or other biological materials that are interactors and may be present in the eluate after affinity chromatography.

In one aspect, the proteins or other biological materials from the extract that were bound to and are eluted from the affinity columns may be resolved for identification by an electrophoresis procedure. Alternatively, this procedure can be omitted and one can proceed directly to identification by mass spectrometry or other analytical methods.

For electrophoresis, polyacrylamide gel electrophoresis (PAGE) on a slab gel may be used. In addition, any of the denaturing or non-denaturing electrophoresis procedures that are standard in the art may be used for this purpose, including gradient gels, capillary electrophoresis, and two-dimensional gels with isoelectric focusing in the first dimension and SDS-PAGE in the second. In certain embodiments, the individual components in the eluent are separated by polyacrylamide gel electrophoresis.

A number of techniques may be used to visualize any protein or other biological material that has been separated by one of the methods described above.

Using electrophoresis, protein bands or spots may be visualized using a staining technique such as Coomassie blue or silver staining, or some other agent that is standard in the art. In certain embodiments, a technique is employed that does not interfere with protein identification by mass spectrometry or use of other analytical methods as described below. Silver staining is often used as it provides a lower detection limit, involves less time for sample preparation and does not lead to protein modifications, at least as compared to other common stains.

Alternatively, autoradiography may be used for visualizing proteins isolated from organisms cultured on media containing a radioactive label, for example $^{35}SO_4^{2-}$ or ^{35}[S]methionine, that is incorporated into the proteins. Radioactive labeling has the advantage of allowing detection and quantitation by scintillation counting of fractions containing binding proteins before polyacrylamide gel electrophoresis. Additionally, the use of radioactively labeled extract allows a distinction to be made between extract proteins

that were retained by the column and proteolytic fragments of the ligand that may be released from the column.

Other labels know to those of skill in the art may also be used for visualization.

Certain proteins and other biological materials separated by one of the foregoing methods will likely be interactors of the protein ligand. In certain instances, such proteins and other materials are those that are derived from the extract (e.g., if a control was used, did not elute from the control column that was not loaded with protein from the extract) and bound to an experimental column that contained protein ligand covalently attached to the solid support, and did not bind to a control column that did not contain any protein ligand. If the separation was achieved by gel electrophoresis, bands of such proteins or other materials may be excised from the stained electrophoretic gel with a clean instrument, usually a scalpel, and further processed for mass spectrometry and other analytical methods as appropriate.

In another aspect, identification of the protein interactor by mass spectrometry is greatly facilitated if the disulfide bonds of the protein are reduced and the free thiols are alkylated after reduction and prior to digestion of the protein with protease. Such a reduction may be performed by treatment of the protein with a reducing agent, for example with dithiothreitol. If the protein is in a gel band after gel electrophoresis, such reduction may occur by treating the band directly. The protein is alkylated by treating with a suitable alkylating agent, for example iodoacetamide.

Prior to analysis by mass spectrometry, the protein may be chemically or enzymatically digested. For protein bands from gels, the protein sample in the gel slice may be subjected to *in-gel* digestion. Shevchenko A. et al., Mass Spectrometric Sequencing of Proteins from Silver Stained Polyacrylamide Gels. Analytical Chemistry 1996, 58, 850-858. One method of digestion is by treatment with the enzyme trypsin, which may be done in-gel. The resulting peptides are extracted from the gel slice into a buffer.

If such a digestion is conducted, the resulting peptide fragments may be purified, for example by use of chromatography. A number of methods are know to those of skill in the art. For example, a solid support that differentially binds the peptides and not the other compounds derived from the gel slice, the protease reaction or the peptide extract may be used. The peptides may be eluted from the solid support into a small volume of a solution that is compatible with mass spectrometry (e.g. 50% acetonitrile/0.1% trifluoroacetic acid)

The preparation of a protein sample from a gel slice that is suitable for mass spectrometry may also be done by an automated procedure.

Mass Spectrometry and Other Analysis

Proteins and other biological materials that interact with the protein ligand may be analyzed by mass spectroscopy, many of which are described in detail below. In addition, the protein and other biological materials may be analyzed by other methods known in the art and are standard in protein chemistry. Such methods may reveal information about the sequence, physical properties, biological activity etc. of the protein and other biological materials.

In one aspect, peptide samples after digestion of the protein interactor may be analyzed by any one of a variety of techniques in mass spectrometry, including, but not limited to matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF), triple quadrupole MS using either electrospray MS, electrospray tandem MS, nano-electrospray MS, or nano-electrospray tandem MS, as well as ion trap or Fourier transform mass spectrometry, or mass spectrometers comprised of components from any one of the above mentioned types (e.g. quadrupole-TOF). This analysis may be performed with any mass spectrometer that has the capability of measuring the peptide masses with adequate mass accuracy, precision, and resolution, as well as the capability of measuring the masses of fragments generated from a specific peptide when analyzed under conditions that induce dissociation of the peptide.

Eluates from the affinity chromatography columns may also be analyzed directly without resolution by electrophoretic methods. In one example, after proteolytic digestion with a protease of the protein of interest, the proteolytic digestion products are applied to a reverse phase column and eluting the peptides from the column directly into a mass spectrometer using an electrospray or nano-electrospray sample introduction interface. For example, peptides may be eluted directly into an ion trap or triple quadrupole mass spectrometer.

Methods that use a MALDI-TOF instrument are, however, more rapid and preferred for high throughput procedures because it takes approximately 30 seconds to analyze a sample by MALDI-TOF in an automated procedure, whereas it takes approximately one hour to introduce samples into the other kinds of instruments via micro-capillary HPLC.

If MALDI-TOF is used to analyze the peptides from the digested interacting protein, the method may yield a high accuracy peptide mass spectrum. Patterson,

Electrophoresis 1995, 16, 1104-14. The peptide masses obtained from MALDI-TOF may be used for correlative database searching of protein or DNA sequence databases. Yates et al., Anal. Biochem. 1993, 214, 397-408. In such a method, the molecular weights of the peptides may be compared with a database of peptides from predicted proteins encoded by the organism's genome, as well as other appropriate databases. This sensitive method is able to characterize proteins that are present at very low concentration, as low as sub-picomole levels in some instances.

This method allows the rapid and accurate mapping of peptide mixtures by measuring the molecular weight of each component. The peptide mixture is generated by sequence-dependent cleavage of the polypeptide backbone by proteolytic enzymes or chemical agents. The peptide map obtained by specific cleavage or digestion, for example with trypsin, results in a unique peptide fingerprint for a given protein. Thus in the case of mass spectrometric mapping, the experimental data are a partial or complete set of molecular weights of peptides resulting from the cleavage (digestion) of the protein. The peptide masses are searched against both in-house proprietary and public databases using a correlative mass matching algorithm. Statistical analysis is performed upon each protein match to determine the validity of the match. Typical constraints include error tolerances within 0.1 Da for monoisotopic peptide masses. Cysteines are alkylated and searched as carboxyamidomethyl modifications. Identified proteins are stored automatically in a relational database with software links to SDS-PAGE images and ligand sequences. Often, even a partial peptide map is specific enough for identification of the protein. If no match is found, a more error-tolerant search can be used, for example using fewer peptides or allowing a larger margin for error. In these cases the tentative identity of the interacting protein should be confirmed by a second method.

This technique is used to assign function to an unknown protein based upon the known function of the interacting protein in the same or a homologous/orthologous organism. Protein- protein interactions are stored in a relational database to create an 'in-silico' network of protein interactions with the predicted effect each protein has upon cellular functions.

The knowledge gained from the relational database is used to select protein targets for further analysis including the immobilization of one or more interacting partners on a solid support and screening a chemical or drug library for compounds that affect the

interaction. The chemicals or drugs are screened for there ability to influence the protein-protein interaction.

Tandem mass spectrometry or post source decay is used for proteins that cannot be identified by peptide-mass matching or to confirm the identity of proteins that are tentatively identified by an error-tolerant peptide mass search, described above. This method combines two consecutive stages of mass analysis to detect secondary fragment ions that are formed from a particular precursor ion. The first stage serves to isolate a particular ion of a particular peptide (polypeptide) of interest based on its m/z. The second stage is used to analyze the product ions formed by spontaneous or induced fragmentation of the selected ion precursor. Interpretation of the resulting spectrum provides limited sequence information for the peptide of interest. However, it is faster to use the masses of the observed peptide fragment ions to search an appropriate protein sequence database and identify the protein as described in Griffin et al, Rapid Commun. Mass. Spectrom. 1995, 9, 1546-51.

Peptide fragment ions are produced primarily by breakage of the amide bonds that join adjacent amino acids. The fragmentation of peptides in mass spectrometry has been well described (Falick et al., J. Am Soc. Mass Spectrom. 1993, 4, 882-893; Biemann, K., Biomed. Environ. Mass Spectrom. 1988, 16, 99-111).

High Throughput and Automation

The methods of the present invention may be conducted in a high throughput fashion and/or by automation.

One non-limiting example of high throughput is repeating a method, or variations of a method, a substantial number of times more quickly than would be possible using standard laboratory techniques. In many instances, the method is used with different samples. By a high throughput method, a single or several individuals may process about 5, 10, 25, 50, 75, 100, 250, 500, 750, 1000, 5000, 10,000 times the number of samples than the same number of individuals would be able to process in the same time period (one, three, seven, 30, 60 90 days).

Automation has been used to achieve high throughput. In regard to automation of the present subject methods, a variety of instrumentation may be used. In general, automation, as used in reference to the subject method, involves having instrumentation complete one or more of the operative steps that must be repeated a multitude of times in performing the method with different samples. Examples of automation include, without

limitation, having instrumentation complete coupling of the protein ligand to the support, adding the extract to a column or other vessel, washings, loading of samples for mass spectroscopy, etc.

There is a range of automation possible for the present invention. For example, the subject methods may be wholly automated or only partially automated. If wholly automated, the method may be completed by the instrumentation without any human intervention after initiating it, other than refilling reagent bottles or monitoring or programming the instrumentation as necessary. In contrast, partial automation of the subject method involves some robotic assistance with the physical steps of the method, such as mixing, washing and the like, but still requires some human intervention other than just refilling reagent bottles or monitoring or programming the instrumentation.

PUBLICATIONS AND OTHER REFERENCES

All publications and patents mentioned herein, including those items listed below, are hereby incorporated by reference in their entirety as if each individual publication or patent was specifically and individually indicated to be incorporated by reference. In case of conflict, the present application, including any definitions herein, will control.

Also incorporated by reference are the following: WO 00/45168, WO 00/79238, WO 00/77712, EP 1047108, EP 1047107, WO 00/72004, WO 00/73787, WO00/67017, WO 00/48004, WO 00/45168, WO 00/45164, U.S.S.N. 09/720,272; PCT/CA99/00640; U.S. Patent Numbers 6,254,833; 6,232,114; 6,229,603; 6,221,612; 6,214,563; 6,200,762; 6,171,780; 6,143,492; 6,124,128; 6,107,477; D428,157; 6,063,338; 6,004,808; 5,985,214; 5,981,200; 5,928,888; 5,910,287; 6,248,550; 6,232,114; 6,229,603; 6,221,612; 6,214,563; 6,200,762; 6,197,928; 6,180,411; 6,171,780; 6,150,176; 6,140,132; 6,124,128; 6,107,066; 6,077,707; 6,066,476; 6,063,338; 6,054,321; 6,054,271; 6,046,925; 6,031,094; 6,008,378; 5,998,204; 5,981,200; 5,955,604; 5,955,453; 5,948,906; 5,932,474; 5,925,558; 5,912,137; 5,910,287; 5,866,548; 5,834,436; 5,777,079; 5,741,657; 5,693,521; 5,661,035; 5,625,048; 5,602,258; 5,552,555; 5,439,797; 5,374,710; 5,296,703; 5,283,433; 5,141,627; 5,134,232; 5,049,673; 4,806,604; 4,689,432; 4,603,209; 6,217,873; 6,174,530; 6,168,784; 6,271,037; 6,228,654; 6,184,344; 6,040,133; 5,910,437; 5,891,993; 5,854,389; 5,792,664; and 6,248,558.

EXEMPLIFICATION

The invention now being generally described, it will be more readily understood by reference to the following examples which are included merely for purposes of illustration

of certain aspects and embodiments of the present invention, and are not intended to limit the invention in any way.

Example 1: Protein SA0005

A protein from the bacterium *Staphylococcus aureus*, labeled SA0005, was chosen for use as the ligand. SA0005 was determined to have high homology to heat shock protein 33, a putative chaperone involved in protein folding.

Production of SA0005

A bioinformatics program (A.L. Delcher, D. Harmon, S. Kasif, O. White, and S.L. Salzberg. Improved microbial gene identification with GLIMMER, Nucleic Acids Research, 1999, 27:, 4636-4641) is used to select the coding sequence of interest from the genome of *S. aureus*. The coding DNA is amplified from purified genomic DNA by using PCR with primers that are identified with a computer program. The PCR primers are selected so as to introduce restriction enzyme cleavage sites at the ends of the DNA (e.g. NdeI and BamH1). The PCR product is purified by gel electrophoresis and directionally cloned into the polylinker of the expression vector pET15b (Novagen, WI) after the polylinker is cut with the same two restriction enzymes. After the ligation reaction, the DNA is transformed into *E. coli* bacteria that will allow the production of the recombinant protein in high yield. The expression vector uses a promoter for the RNA polymerase of bacteriophage T7, and the strain of *E. coli* is able to produce T7 RNA polymerase when isopropyl-β-D-thiogalactoside (IPTG) is added to the growth medium. The sequence of the cloning site is such as to add polyhistidine, followed by a cleavage site for the enzyme thrombin, to the amino-terminal of the recombinant heterologous protein. Bacteria containing the recombinant plasmid are selected for by antibiotic resistance, indicating they have acquired the plasmid, and identified either by using PCR or another method to analyze their DNA or by using SDS-PAGE or mass spectrometry to identify clones that produce the desired protein in large amounts.

A clone that produces the desired recombinant heterologous protein in large amounts is grown in Luria broth or another medium. IPTG is added when the culture has reached an appropriate cell density and then the culture is incubated overnight at 15°C, harvested by centrifugation at 5000 rpm for 15 minutes, and broken by sonication. The extract is clarified by centrifugation at 15000 rpm for 30 minutes. Nucleic acid is removed from the clarified extract by passing the extract through a DE52 column in a buffer containing 500 mM NaCl. The recombinant protein is then bound to a nickel column and

eluted with buffer containing imidazole. After the imidazole is removed from the preparation by dialysis, the tag is removed from the protein by digestion with thrombin and the mixture is passed through another nickel column. The recombinant heterologous protein without the polyhistidine tag flows through the second nickel column, now highly purified and ready for use in affinity chromatography.

Staphylococcus aureus extract preparation:

A *Staphylococcus aureus* extract is prepared from cell pellets using nuclease and lysostaphin digestion followed by sonication. A *Staphylococcus aureus* cell pellet (12g) is suspended in 12 ml of 20 mM Hepes pH 7.5, 150 mM NaCl, 10% glycerol, 10 mM $MgSO_4$, 10 mM $CaCl_2$, 1 mM DTT, 1 mM PMSF, 1 mM benzamidine, 1000 units of lysostaphin, 0.5 mg RNAse A, 750 units micrococcal nuclease, and 375 units DNAse 1. The cell suspension is incubated at 37°C for 30 minutes, cooled to 4°C, and is made up to a final concentration of 1 mM EDTA and 500 mM NaCl. The lysate is sonicated on ice using three bursts of 20 seconds each. The lysate is centrifuged at 20 000 rpm for 1 hr in a Ti70 fixed angle Beckman rotor. The supernatant is removed and dialyzed overnight in a 10 000 Mr dialysis membrane against ACB (20 mM Hepes pH 7.5, 10 % glycerol, 1 mM DTT, and 1 mM EDTA) containing 100 mM NaCl, 1mM benzamidine, and 1 mM PMSF. The dialyzed protein extract is removed from the dialysis tubing and frozen in one ml aliquots at –70°C.

Preparation of Affinity Column

A series of solutions of the ligand (SA0005) is prepared so as to give final amounts of 0, 0.1, 0.5, 1.0, and 2.0 mg of ligand per ml of resin volume. Assuming that the stock solution of ligand has a concentration of 3.5 mg/ml the following samples are prepared in labeled silanized microcentifuge tubes:

ligand conc. on resin	0	0.1	0.5	1	2
volume of resin (µl)	100	100	100	100	100
Protein (µg)	0	10	50	100	200
protein (µl)	0.0	2.9	14.5	28.9	57.8
ACB buffer (µl)	300	297.1	285.5	271.1	242.2

A slurry of Affigel 10 is prepared and 1 ml of slurry is removed (enough for six 100-ml aliquots of resin). Using a glass frit Buchner funnel, the resin is washed sequentially with three 10 ml portions each of ice-cold isopropanol, distilled H_2O, and ACB containing 1 M NaCl. The resin is completely drained of buffer, but not dried. Into six clean silanized microcentrifuge tubes is added 100 mg of the Affigel 10. The buffer containing the ligand concentration series, as shown in the table, is added to the tubes containing Affigel 10 and mixed gently. The tubes containing the coupling reactions are places on a rotator at 4°C overnight. After coupling, the Affigel 10 resin is centrifuged at 2000 rpm for 1 minute at 4°C, or alternatively, the beads are allowed to settle under gravity. The beads are isolated by removing the supernatant solution which is saved for later analysis to evaluate the coupling efficiency.

To the Affigel 10 is added 300 µl of ACB containing 100 mM NaCl and 80 mM ethanolamine. The Affigel 10 is resuspended and rotated for 2 hours at 4°C. The remaining reactive groups react with the ethanolamine. The Affigel 10 resin is centrifuged at 2000 rpm for 1 minute at 4°C, or the beads are allowed to settle under gravity. The supernatant is removed and discarded. As an option, add 300 µl of ACB containing 100 mM NaCl and 1 mg/ml of bovine serum albumin, resuspend the beads, and rotate for 2 hours. The Affigel 10 resin is centrifuged at 2000 rpm for 1 minute at 4°C, or allowed to settle under gravity, and the supernatant is removed and discarded. The resin is resuspended in 300 µl of ACB containing 1 M NaCl. This step is repeated 3 times to wash away the free bovine serum albumin from the resin. The supernatant is removed and the resin is resuspended with 100 µl of ACB containing 100 mM NaCl.

The micro-columns are prepared by using forceps to bend the ends of P200 pipette tips. To the pipette tips is added 10µl of glass beads and 80 µl of a 50% slurry of the Affigel 10 resin containing the covalently attached ligand protein. The columns are allowed to drain on ice in a 1.5 ml microcentrifuge tube and are washed with 10 column volumes (400 µl) of ACB containing 100 mM NaCl.

Affinity chromatography

Ten column volumes of the *S. aureus* extract is added to each micro-column and the flow-throughs of the columns are removed when approximately 50 - 100 µl accumulates. Each column is washed in the same manner with 5 column volumes of ACB containing 100 mM NaCl. This washing is repeated once. Each column is washed with 5 column volumes

of ACB containing 100 mM NaCl and 0.1% Triton X-100. The columns are eluted sequentially with 4 column volumes of ACB containing 1M NaCl and 4 column volumes of 1% sodium dodecyl sulfate into clean microcentrifuge tubes. To each eluted fraction is added one-tenth volume of 10-fold concentrated loading buffer for SDS-PAGE.

Resolution of the eluted proteins and detection of bound proteins

The components of the eluted samples are resolved on SDS-polyacrylamide gels containing 13.8% polyacrylamide using the Laemmli buffer system.

After the electrophoresis procedure is complete, the gel is stained in a clean glass tray. Using 500 ml of each rinse solution, the gel is treated sequentially with 1) 50% methanol, 10% acetic acid overnight or for at least two hours to fix the gel. Repeat once for 20 minutes, 2) 20% ethanol for 10 minutes, 3) distilled water for 10 minutes, 4) sodium thiosulfate (0.2 g/liter) for 1 minute to reduce the gel, 5) water, twice for 20 seconds each wash, 6) silver nitrate (2.0 g/liter) for 30 minutes, and 7) water for 20 seconds. The gel is washed once with developing solution (50 to 75 ml) for 30 seconds, and is developed to the desired intensity, until the band is visible (a light to dark brown). The developing solution contains sodium carbonate (30 g/liter), formaldehyde (1.4 ml of 37% solution/liter), and sodium thiosulfate (10 mg/liter). Once the desired stain intensity has been reached, the developing solution is removed quickly. The reaction is stopped by adding a 1% acetic acid solution and incubating for a minimum of 20 minutes. The gel is rinsed with 1% acetic acid.

The gel is shown in Figure 1. One interacting protein is apparent from the 1% SDS eluates.

The bands containing the interacting protein are excised with a clean scalpel. The gel volume is kept to a minimum by cutting as close to the band as possible. The gel slice is placed into a clean 0.5 ml microcentrifuge tube. To the gel slices is added 10 to 20 µl of 1% acetic acid. The sample can be stored frozen at -70°C for an extended period of time.

Sample Preparation for Mass Spectrometry

The gel slices are cut into 1 mm cubes and 10 to 20 µl of 1% acetic acid is added. The gel particles are washed with 100 - 150 µl of HPLC grade water (5 minutes with occasional mixing), briefly centrifuged and the liquid is removed. Acetonitrile (~200 µl, approximately 3 to 4 times the volume of the gel particles) is added followed by incubation at room temperature for 10 to 15 minutes with occasional mixing. A second acetonitrile

wash may be required to completely shrink the gel particles. The sample is briefly centrifuged and all the liquid is removed.

The protein in the gel particles is reduced by covering the gel slices with 100 mM ammonium bicarbonate containing 10 mM dithiothreitol and incubating at 50°C for 30 minutes.

Briefly centrifuge and remove all the liquid. Acetonitrile is added to shrink the gel particles and the excess liquid is removed. The protein in the gel particles is alkylated by covering the gel particles with 100 mM ammonium bicarbonate containing 55 mM iodoacetamide and incubating for 20 minutes at room temperature in the dark. The sample is briefly centrifuged and all the liquid is removed. The gel particles are washed with 150 to 200 µl of 100 mM ammonium bicarbonate for 15 minutes with occasional mixing. The sample is briefly centrifuged and all the liquid is removed. Acetonitrile is added to shrink the gel particles and the excess liquid is removed. The sample is briefly centrifuged and all the liquid is removed. The gel particles are dried using a centrifugal vacuum concentrator for 1 minute.

To digest the interacting protein, the gel particles are rehydrated in digestion buffer containing trypsin (50 mM ammonium bicarbonate, 5 mM $CaCl_2$, and 12.5 ng/µl trypsin) on ice for 30 to 45 minutes (after 20 minutes incubation more trypsin solution is added). The excess trypsin solution is removed and 10 to 15 µl digestion buffer without trypsin is added to ensure the gel particles remain hydrated during digestion. The samples are incubated at 37°C overnight.

The samples are briefly centrifuged and all the liquid is transferred to a clean microcentrifuge tube (0.5 ml)(step 1). To the gel particles is added 100 µl of 100 mM ammonium bicarbonate and the peptides are extracted by shaking at 37°C in an orbital shaker for 30 minutes followed by centrifugation. The liquid (step 2) is pooled with the liquid from step 1. A second portion of 100 µl of 100 mM ammonium bicarbonate is added to the gel particles and the peptides are extracted a second time by shaking at 37°C in an orbital shaker for 30 minutes followed by centrifugation. The liquid is pooled with the liquid from steps 1 and 2.

Purification of the tryptic peptides

Bulk C18 reverse phase resin is washed several times with methanol and with 65% acetonitrile prior to use and a 5:1 slurry is prepared with 65% acetonitrile/1% acetic acid.

Five µl of the C18 slurry are added to the extracted peptides and shaken for 30 minutes at 37°C. The supernatant is removed and 150 µl of 2% acetonitrile/1% acetic acid are added and shaken for 5 to 15 minutes at 37°C. All of the supernatant is removed and 10 to 15 µl of 65% acetonitrile / 1% acetic acid are added. The sample is vortexed briefly and incubated for 5 minutes with occasional mixing. The sample is centrifuged and the supernatant is removed to a fresh tube for analysis by mass spectrometry.

Mass spectrometric analysis

Analytical samples containing tryptic peptides are subjected to Matrix Assisted Laser Desorption/Ionization Time Of Flight (MALDI-TOF) mass spectrometry. Samples are initially mixed with an equal volume of organic solvent containing a compound (matrix) that ionizes peptides upon excitation by a laser pulse. The matrix could be one of α-cyano-4-hydroxy-*trans*-cinnamic acid, sinnipinic acid, or 2,5-dihydroxybenzoic acid. The mixture of the sample and matrix is allowed to dry on a sample stage and introduced into the mass spectrometer. Specifically, 0.5 µl matrix solution containing 20 mg/ml α-cyano-4-hydroxy-*trans*-cinnamic acid in 50% acetonitrile/1% acetic acid is mixed with 0.5 µl sample and applied to a well of a multi-sample MALDI-TOF plate. Analysis of the peptides in the mass spectrometer is carried out using delayed extraction and an ion reflector to ensure high resolution of peptides. The instrument is initially calibrated using the autohydrolysis peaks generated by trypsin, but the method is not dependent upon trypsin and any protease having a defined cleavage specificity may be used.

Tryptic peptide masses are searched against both in-house proprietary and public databases using a correlative mass matching algorithm. Twenty peptide masses were used in the search. Statistical analysis is performed upon each protein match to determine the validity of the match. Typical constraints include error tolerances within 0.1 Da for monoisotopic peptide masses. Cysteines are alkylated and are searched as carboxyamidomethyl modifications. Identified proteins are stored automatically in a relational database with software links to SDS-PAGE images and ligand sequences. The tryptic peptide mass spectrum is shown in Figure 2. The closest protein match from the correlative search and the probability of a correct match for the five closest protein matches are shown in Table 1.

Table 1: Results of correlative database searching of 20 peptide masses.

Rank	Probability	Name

1	1.0e+00	EF-Tu
2	1.0e-17	
3	7.7e-18	
4	1.4e-18	
5	2.0e-19	

One interacting protein was discovered and identified as a truncated form of EF-Tu, whose intact form is a key factor involved in protein biosynthesis. This form of EF-Tu is novel. It is most likely made by intracellular proteolysis from intact EF-Tu. It could be involved in protein synthesis in *S. aureus* or could have some other function. The chaperone, if it is one, could be involved in the folding of the EF-Tu fragment or in its assembly with some other protein.

Examples 2 -5 are performed using the procedures of Example 1.

Example 2: protein SA0146

A protein from the bacterium *Staphylococcus aureus*, labeled SA0146, was chosen for use as the ligand. SA0146 was found to be a homolog of the *B. subtilis* cell division initiation protein, DIV IVA, which is involved in septum formation.

Table 2: Results of correlative database searching of 14 peptide masses .

Rank	Probability	Name
1	1.0e+00	conserved protein of unknown function
2	1.7e-11	
3	4.0e-12	
4	8.5e-13	
5	8.2e-13	

The interacting protein was found to be a conserved protein of unknown function. The data suggests that the interacting conserved protein is also involved in cell division. It could be a good drug target because cell division is an essential process.

Example 3: protein SA0203

An unknown protein from the bacterium *Staphylococcus aureus*, labeled SA0203, was chosen for use as the ligand. The function of SA0203 is unknown.

Table 3: Results of correlative database searching of 15 peptide masses.

Rank	Probability	Name
1	1.0e+00	peptide chain release factor 3
2	2.4e-07	
3	2.5e-08	
4	1.1e-08	
5	8.1e-09	

The interacting protein was found to be a homologue of peptide chain release factor 3. Its interaction with peptide chain release factor 3 suggests that it is involved in the termination stage of protein synthesis. It could potentially be a good drug target because many antibiotics inhibit protein synthesis.

Example 4: protein SA0276

A protein from the bacterium *Staphylococcus aureus*, labeled SA0276, was chosen for use as the ligand. Because of its high homology to other bacterial homologues, SA0276 was labeled a putative phenylalanine tRNA synthetase subunit, although only part of its sequence is a good match to enzymes of that type in other species.

Table 4: Identification of Interactor 1, results of correlative database searching of 29 peptide masses.

Rank	Probability	Name
1	1.0e+00	glutamyl-tRNA Gln amidotransferase subunit B
2	7.7e-22	
3	6.3e-22	
4	5.1e-23	
5	6.4e-24	

Table 5: Identification of Interactor 2, results of correlative database searching of 23 peptide masses.

Rank	Probability	Name
1	1.0e+00	glutamyl-tRNA Gln amidotransferase subunit A
2	1.9e-13	
3	1.3e-14	
4	3.4e-15	
5	1.7e-15	

Two interacting proteins were discovered and identified as homologues of glutamyl-tRNA Gln amidotransferase subunits A and B.

In *S. aureus* and perhaps other organisms, SA0276 may have an additional function in which it interacts with a portion of glutamyl tRNA and acts as a cofactor for glutamyl-tRNA glutamine amidotransferase. If so, that might have a vital function outside of charging phenylalanine tRNA, and chemicals that inhibit that activity could be good antibiotics.

Example 5: protein SA0526

A protein from the bacterium *Staphylococcus aureus*, labeled SA0526, was chosen for use as the ligand. SA0526 was determined to be a homologue of EF-Ts, a protein synthesis elongation factor that is conserved in all bacteria.

Table 6: Results of correlative database searching of 14 peptide masses.

Rank	Probability	Name
1	1.0e+00	EF-Tu
2	4.6e-12	
3	3.7e-12	
4	1.6e-13	
5	1.4e-13	

The interacting protein was found to be a homologue of EF-Tu. The interaction of EF-Tu with EF-Ts, which is confirmed in this experiment, has been known for more than 30 years.

Example 6: protein SA0808

A protein from the bacterium *Staphylococcus aureus*, labeled SA0808, was chosen for use as the ligand. SA0808 was determined to be homologous to menaquinone biosynthesis methyltransferase, an enzyme involved in the last step in the synthesis of menaquinone (vitamin K).

SA0808 was prepared in a manner analogous to example 1.

S. aureus extract preparation:

A *S. aureus* cell pellet (~12g) is suspended in 20 ml of lysis buffer (20 mM Hepes pH 7.5, 500 mM NaCl, 10% glycerol, 10 mM $MgSO_4$, 10 mM $CaCl_2$, 1 mM DTT, 1mM EDTA, 1 mM PMSF, 1 mM benzamidine). The nucleases Rnase A (40 µg/ml final) and micrococcal nuclease (75 units/mL) are added. The cells are lysed with 10 pulses of 30 sec. between 90 sec. pauses using the Bead-Beater apparatus (Biospec Products Inc.). The outer chamber of the apparatus is filled with ice and the inner chamber with a 50/50 mixture of cells and zirconia beads (0.1 mm diameter). The lysate is separated from the zirconia beads using a standard chromatography column and peristalic pump. The lysate is centrifuged at 20000 rpm (48000 x g) in Oak Ridge tubes (50 mL capacity) in a Beckman JA25.50 rotor. The extract is dialyzed against 1 L of 0.1 M ACB (20 mM Hepes pH 7.5, 100 mM NaCl, 10% glycerol, 10 mM $MgSO_4$, 10 mM $CaCl_2$, 1 mM DTT, 1mM EDTA, 1 mM PMSF, 1 mM benzamidine) overnight at 4°C in a dialysis membrane (Spectrum Labs, 10 kDa size exclusion). The extract is removed from the dialysis membrane and stored in 1 mL aliquots at -80°C.

Preparation of Affinity Column

A series of solutions of the ligand (SA0808) is prepared so as to give final amounts of 0, 0.1, 0.5, 1.0, and 2.0 mg of ligand per ml of resin volume. Assuming that the stock solution of ligand has a concentration of 3.5 mg/ml the following samples are prepared in labeled silanized microcentifuge tubes:

ligand conc. on resin	0	0.1	0.5	1	2
volume of resin (µl)	100	100	100	100	100
Protein (µg)	0	10	50	100	200
protein (µl)	0.0	2.9	14.5	28.9	57.8
ACB buffer (µl)	300	297.1	285.5	271.1	242.2

A slurry of Affigel 10 is prepared and 1 ml of slurry is removed (enough for six 100-ml aliquots of resin). Using a glass frit Buchner funnel, the resin is washed sequentially with three 10 ml portions each of ice-cold isopropanol, distilled H_2O, and ACB containing 1 M NaCl. The resin is completely drained of buffer, but not dried. Into six clean silanized microcentrifuge tubes is added 100 mg of the Affigel 10. The buffer containing the ligand concentration series, as shown in the table, is added to the tubes containing Affigel 10 and gently mixed to suspend the resin. The tubes containing the coupling reactions are places on a rotator at 4°C overnight. After coupling, the Affigel 10 resin is centrifuged at 2000 rpm for 1 minute at 4°C. The beads are isolated by removing the supernatant solution. The supernatant of the 2 mg/ml reaction is saved for later analysis to evaluate the coupling efficiency.

To remove any free ligand, the resin is resuspended with 1 M ACB, centrifuged at 2000 rpm, and the supernatant is removed. This is repeated twice more. The resin is resuspended with 100 µL of 0.1 M ACB.

The micro-columns are prepared by using forceps to bend the ends of P200 pipette tips. To the pipette tips is added 10 µl of glass beads and 80 µl of a 50% slurry of the Affigel 10 resin containing the covalently attached ligand protein. The micro-columns are allowed to drain on ice in a 1.5 ml microcentrifuge tube. The micro-columns are adjusted to 40 µl of resin (50 µl mark on tip) and are washed with 5 column volumes of ACB containing 100 mM NaCl.

Affinity chromatography

The extract is centrifuged in a microcentrifuge tube at 15000 rpm for 15 minutes at 4°C. The supernatant is removed to a fresh microcentrifuge tube and diluted to 5 mg protein/ml ACB containing 100 mM NaCl.

Five column volumes of the *S. aureus* extract is added to each micro-column and the flow-throughs of the columns are removed when approximately 50 - 100 ml accumulates.

Each column is washed in the same manner with 5 column volumes of ACB containing 100 mM NaCl. This washing is repeated once. Each column is washed with 5 column volumes of ACB containing 100 mM NaCl and 0.1% Triton X-100. The columns are eluted sequentially with 4 column volumes of 1% sodium dodecyl sulfate into clean microcentrifuge tubes. To each eluted fraction is added one-tenth volume of 10-fold concentrated gel loading buffer.

Resolution of the eluted proteins and detection of bound proteins

The components of the eluted samples are resolved on polyacrylamide gels (no SDS is present in the gel, with 0.1% present in the gel running buffer) containing 13.8% polyacrylamide.

The gels are stained by silver staining using a mass spectrometry compatible protocol, as in Example 1. The gel is shown in Figure 11.

The bands of interest are excised with a clean scalpel. The gel volume is kept to a minimum by cutting as close to the band as possible. The gel slice is placed into a clean 0.5 ml microcentrifuge tube. To the gel slices is added 10 to 20 µl of 1% acetic acid. The sample can be stored frozen at -70°C for an extended period of time.

Sample Preparation for Mass Spectrometry

The interacting proteins in the excised bands are digested with trypsin, and the resulting peptides are purified according to the procedures of Example 1.

Mass spectrometric analysis

The tryptic peptides are analyzed using MALDI-TOF mass spectrometry according to the procedures of Example 1. The tryptic peptide masses are searched against both in-house proprietary and public databases using a correlative mass matching algorithm. Statistical analysis is performed upon each protein match to determine the validity of the match. Typical constraints include error tolerances within 0.1 Da for monoisotopic peptide masses. Cysteines are alkylated and are searched as carboxyamidomethyl modifications. Identified proteins are stored automatically in a relational database with software links to SDS-PAGE images and ligand sequences. The tryptic peptide mass spectra for the four interacting proteins are shown in Figures 12a and 12b. The closest protein match from each correlative search and the probability of a correct match for the five closest protein matches are shown in Tables 7-10.

Table 7: Identification of Interactor 1, results of correlative database searching of 27 peptide masses.

Rank	Probability	Name
1	1.0e+00	elongation factor G
2	1.2e-28	
3	2.4e-30	
4	1.7e-30	

5	1.6e-30	

Table 8: Identification of Interactor 2, results of correlative database searching of 21 peptide masses.

Rank	Probability	Name
1	1.0e+00	trigger factor (prolyl isomerase)
2	2.9e-10	
3	1.4e-10	
4	7.1e-11	
5	5.0e-11	

Table 9: Identification of Interactor 3, results of correlative database searching of 19 peptide masses.

Rank	Probability	Name
1	1.0e+00	formate-tertrahydrofolate ligase
2	1.9e-07	
3	7.3e-08	
4	2.9e-08	
5	1.8e-08	

Table 10: Identification of Interactor 4, results of correlative database searching of 29 peptide masses.

Rank	Probability	Name
1	1.0e+00	EF-Tu
2	1.3e-27	
3	7.0e-28	
4	1.0e-28	
5	3.2e-29	

Four interacting proteins are discovered and identified by MALDI-TOF mass spectrometry and correlative database searching as homologues of elongation factor G, trigger factor (prolyl isomerase), formate-tetrahydrofolate ligase, and EF-Tu.

SA0808 is homologous to an enzyme involved in the last step in the synthesis of menaquinone (vitamin K). Its involvement in single carbon transfer as a methyltransferase could explain its interaction with formate-tetrahydrofolate ligase, an enzyme involved in one-carbon metabolism, but the exact connection is obscure. SA0808 also interacts with trigger factor, which is a prolyl isomerase. The prolyl isomerase could be involved in the proper folding of SA0808 or could have some other role in its activity. There is genetic evidence for the possible involvement of the homologue of SA0808 of *B. subtilis* in spore germination, which involves the restart of a variety of metabolic processes, including protein synthesis. That could suggest that SA0808 has a previously unsuspected function in which it interacts with and perhaps modifies the protein synthesis factors EF-Tu and EF-G in order to control their activities. Interfering with this interaction could be a way to control the germination of bacteria.

Examples 7-10 are performed using the procedures of Example 6.

Example 7: protein SA0989

A protein from the bacterium *Staphylococcus aureus*, labeled SA0989, is chosen for use as the ligand. SA0989 was determined to be homologous to 3-methyl-2-oxobutanoate dehydrogenase.

Table 12: Identification of Interactor 1; Results of correlative database searching of 24 peptide masses.

Rank	Probability	Name
1	1.0e+00	trigger factor (prolyl isomerase)
2	2.4e-20	
3	1.9e-21	
4	7.7e-22	
5	7.7e-22	

Table 13: Identification of Interactor 3; Results of correlative database searching of 13 peptide masses.

Rank	Probability	Name
1	1.0e+00	enolase
2	1.4e-07	
3	7.5e-08	

4	6.0e-08	
5	7.5e-09	

Three interacting proteins are discovered Two are identified by MALDI-TOF mass spectrometry as homologues of trigger factor (prolyl isomerase) and enolase. The third is unidentified.

SA0989 is probably a branched chain α-ketoacid dehydrogenase involved in the second step in the synthesis of branched chain amino acids. Trigger factor is a prolyl isomerase which could be involved in the folding of SA0989. SA0989 also interacts with another protein that has not yet been positively identified and with enolase. Although the interaction with enolase could have some significance that we do not appreciate, enolase has been found to bind to at least 20 of the proteins of *S. aureus*. Although it is possible that enolase has a chaperone-like function for many other proteins, it is also possible that enolase is a protein that interacts with many proteins in a fashion that is not biologically important.

Example 8: protein SA1094

A protein from the bacterium *Staphylococcus aureus*, labeled SA1094, is chosen for use as the ligand. SA1094 is a protein of heretofore unknown function.

Table 14: Results of correlative database searching of 29 peptide masses.

Rank	Probability	Name
1	1.0e+00	putative peptidase
2	1.7e-19	
3	1.7e-21	
4	3.9e-22	
5	2.1e-22	

One interacting protein is discovered. The interactor is found to be a homologue of a putative peptidase.

The interaction of SA1094 with a putative peptidase (based on homologues in other organisms) suggests that SA1094 is likely to be involved in peptide metabolism.

Example 9: protein SA1185

A protein from the bacterium *Staphylococcus aureus*, labeled SA1185, is chosen for use as the ligand. SA1185 is a protein of heretofore unknown function.

Table 15: Identification of Interactor 1; Results of correlative database searching of 39 peptide masses.

Rank	Probability	Name
1	1.0e+00	glucose-6-phosphate isomerase
2	3.1e-36	
3	2.6e-36	
4	1.1e-36	
5	5.6e-37	

Table 16: Identification of Interactor 2; Results of correlative database searching of 35 peptide masses.

Rank	Probability	Name
1	1.0e+00	cysteine synthetase
2	7.9e-40	
3	1.3e-41	
4	5.1e-43	
5	2.1e-43	

Two interacting proteins are discovered. The identities of the interactors are determined by MALDI-TOF mass spectrometry as homologues of glucose-6-phosphate isomerase and cysteine synthetase.

SA1185 interacts with two enzymes of widely differing functions, glucose-6-phosphate isomerase involved in glucose metabolism and cysteine synthetase involved in the last step in cysteine biosynthesis. SA1185 could be involved in controlling the activities or localizations of both enzymes.

Example 10: protein SA1203

A protein from the bacterium *Staphylococcus aureus*, labeled SA1203, is chosen for use as the ligand. SA1203 is a protein of heretofore unknown function.

Table 12: Results of correlative database searching of 21 peptide masses.

Rank	Probability	Name
1	1.0e+00	NADH dehydrogenase.
2	1.9e-14	
3	3.6e-16	
4	4.2e-17	
5	2.9e-17	

One interacting protein is discovered. The interacting protein is a homologue of NADH dehydrogenase.

SA1203's specific interaction with the respiratory enzyme NADH dehydrogenase suggests it could be involved in respiration and controlling the activity or membrane versus cytosolic location of that enzyme.

EQUIVALENTS

While specific embodiments of the subject invention have been discussed, the above specification is illustrative and not restrictive. Many variations of the invention will become apparent to those skilled in the art upon review of this specification. The appended claims are not intended to claim all such embodiments and variations, and the full scope of the invention should be determined by reference to the claims, along with their full scope of equivalents, and the specification, along with such variations.

1. A method for the identification of an interacting protein, the method comprising:
   a) subjecting an extract to protein-affinity chromatography on two or more columns, the columns having a protein ligand in varying concentrations immobilized to a matrix, and eluting bound components of the extract from the columns;
   b) separating the components to isolate an interacting protein;
   c) analyzing the interacting protein by mass spectrometry to identify the interacting protein.
2. The method of claim 1, wherein the columns are micro-columns.
3. The method of claim 2, wherein multiple micro-columns are arranged into an array format.
4. The method of claim 1, wherein the columns are not blocked after immobilizing the ligand to the matrix.
5. The method of claim 1, wherein the protein ligand is immobilized to the matrix after the matrix has been packed into the column.
6. The method of claim 1, wherein the separation is a gel-separation.
7. The method of claim 6, wherein said gel-separation is a polyacrylamide gel electrophoresis.
8. The method of claim 7, wherein said polyacrylamide gel contains SDS.
9. The method of claim 1, wherein said protein ligand is covalently bound to the matrix.
10. The method of claim 1, wherein said mass spectrometry is MALDI-TOF mass spectrometry.
11. The method of claim 1, wherein the bound components of the extract are eluted with a protein denaturant.
12. The method of claim 1, wherein the protein-affinity chromatography is an automated process.
13. The method of claim 12, wherein the automated process includes procedures for preparing the columns and performing the affinity chromatography.
14. The method of claim 13, wherein the automated process includes procedures for packing the columns, coupling the protein ligand to the matrix, loading an extract onto the columns, washing the columns and eluting bound components from the columns.

15. The method of claim 1, wherein the protein ligand is at least 90% pure.

16. The method of claim 1, wherein the protein ligand is a fusion protein.

17. The method of claim 16, wherein the fusion protein comprises an affinity tag which may be used to couple the protein ligand onto the matrix.

18. The method of claim 1, wherein the concentration of the protein ligand bound to the matrix in at least one of the columns is at least 10-fold higher than the K_d of the interaction between the protein ligand and the interacting protein.

19. The method of claim 1, wherein the concentration of the protein ligand bound to the matrix is from 0 to about 2 milligrams of ligand per milliliter of matrix for all of the columns.

20. The method of claim 1, wherein the extract is derived from a tissue, cultured cell line, purified cellular organelle, or bodily fluid.

21. The method of claim 1, wherein the extract is a whole cell extract or a fractionated extract.

22. A method for the identification of an interacting protein, said method comprising:

a) subjecting a cellular extract or extracellular fluid to protein-affinity chromatography on two or more columns, said columns having a protein ligand coupled to the matrix in varying concentrations, and eluting bound components of said extract from said columns;

b) gel-separating said components to isolate an interacting protein; wherein the interacting protein is observed to vary in amount in direct relation to the concentration of coupled protein ligand;

c) digestion of said interacting protein to give corresponding peptides; and

d) analyzing said peptides by MALDI-TOF mass spectrometry or post source decay to determine the peptide masses.

23. The method of claim 22, wherein said columns are micro-columns.

24. The method of claim 22, wherein the columns are not blocked after coupling the ligand to the matrix.

25. The method of claim 22, wherein the protein ligand is coupled to the matrix after the matrix has been packed into the column.

26. The method of claim 22, wherein said gel-separation is a polyacrylamide gel electrophoresis.

27. The method of claim 26, wherein said polyacrylamide gel contains SDS.

28. The method of claim 22, wherein said protein ligand is covalently bound to the matrix.

29. The method of claim 22, wherein the identities of the interacting protein partners are entered into a relational database.

30. The method of claim 22, wherein the bound components of the extract are eluted with a protein denaturant.

31. The method of claim 22, wherein the protein-affinity chromatography is an automated process.

32. The method of claim 31, wherein the automated process includes procedures for preparing the columns and performing the affinity chromatography.

33. The method of claim 32, wherein the automated process includes procedures for packing the columns, coupling the protein ligand to the matrix, loading an extract onto the columns, washing the columns and eluting bound components from the columns.

34. The method of claim 22, further comprising correlative database searching with said peptide or peptide fragment masses, whereby the interacting protein is identified.

Figure 1. Interactions with SA0005

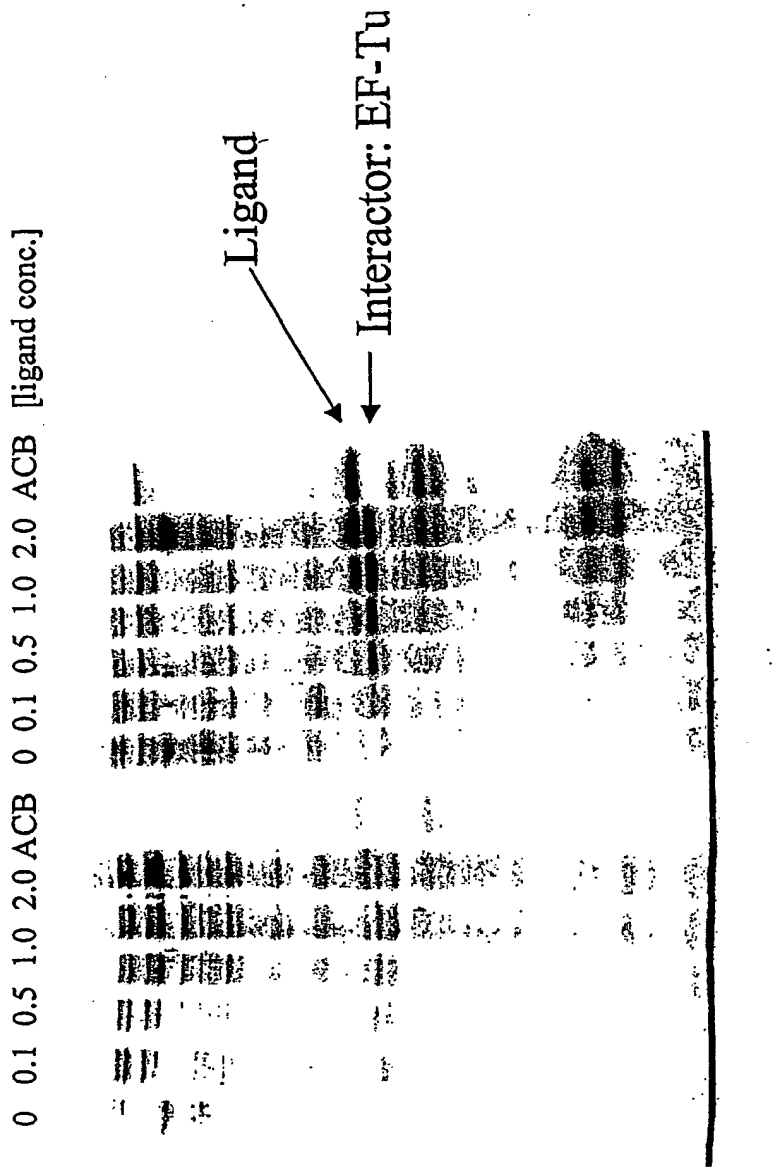


Figure 2: Interactions with SA0005

MALDI-TOF Mass Spectrum of Interactor (EF-Tu)

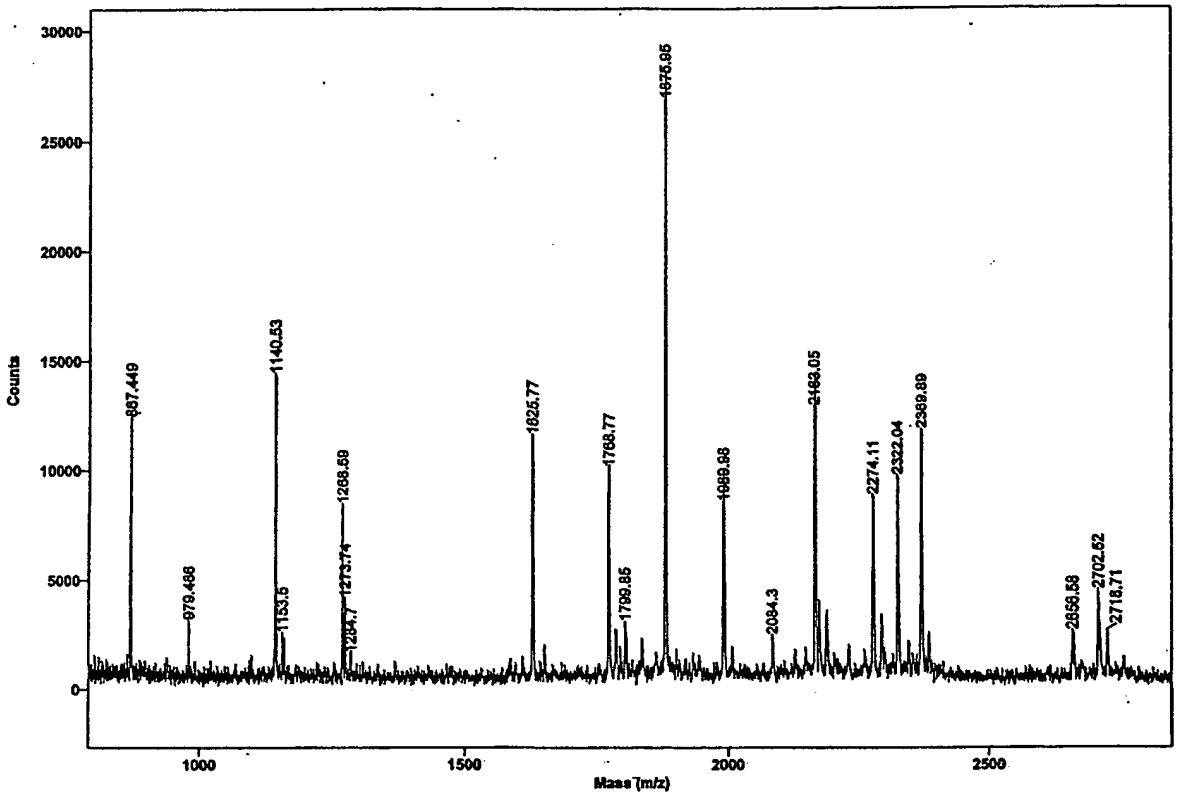


Figure 3. Interactions with SA0146

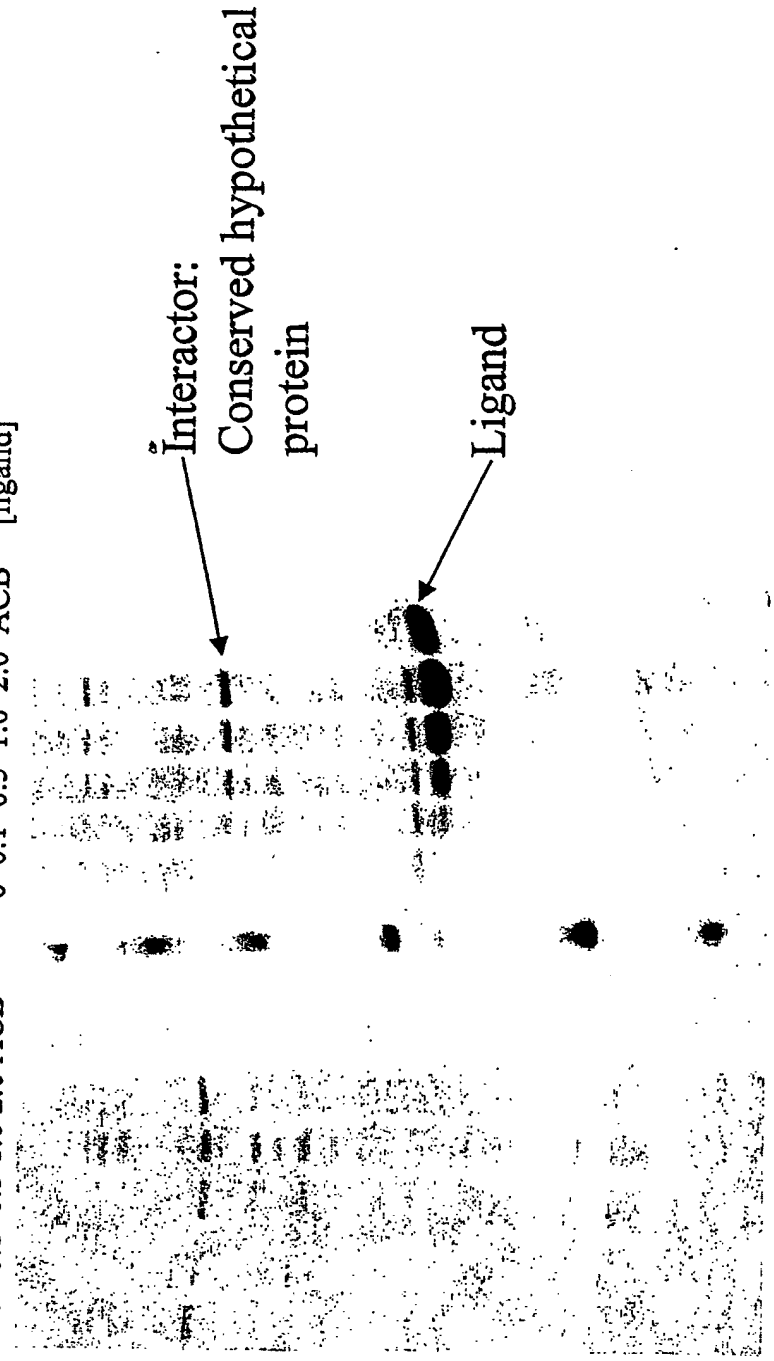


Figure 4: Interactions with SA0146

MALDI-TOF Mass Spectrum of Interactor (conserved hypothetical protein)

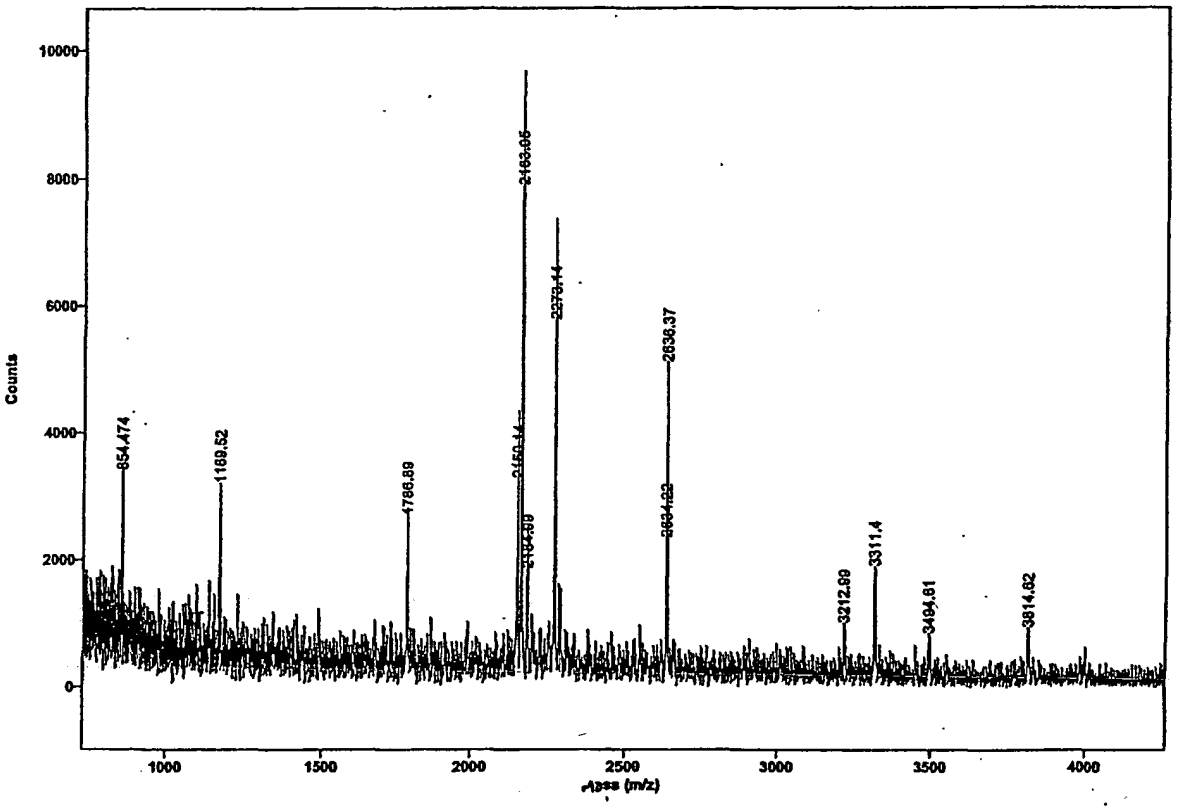


Figure 5: Interactions with SA0203

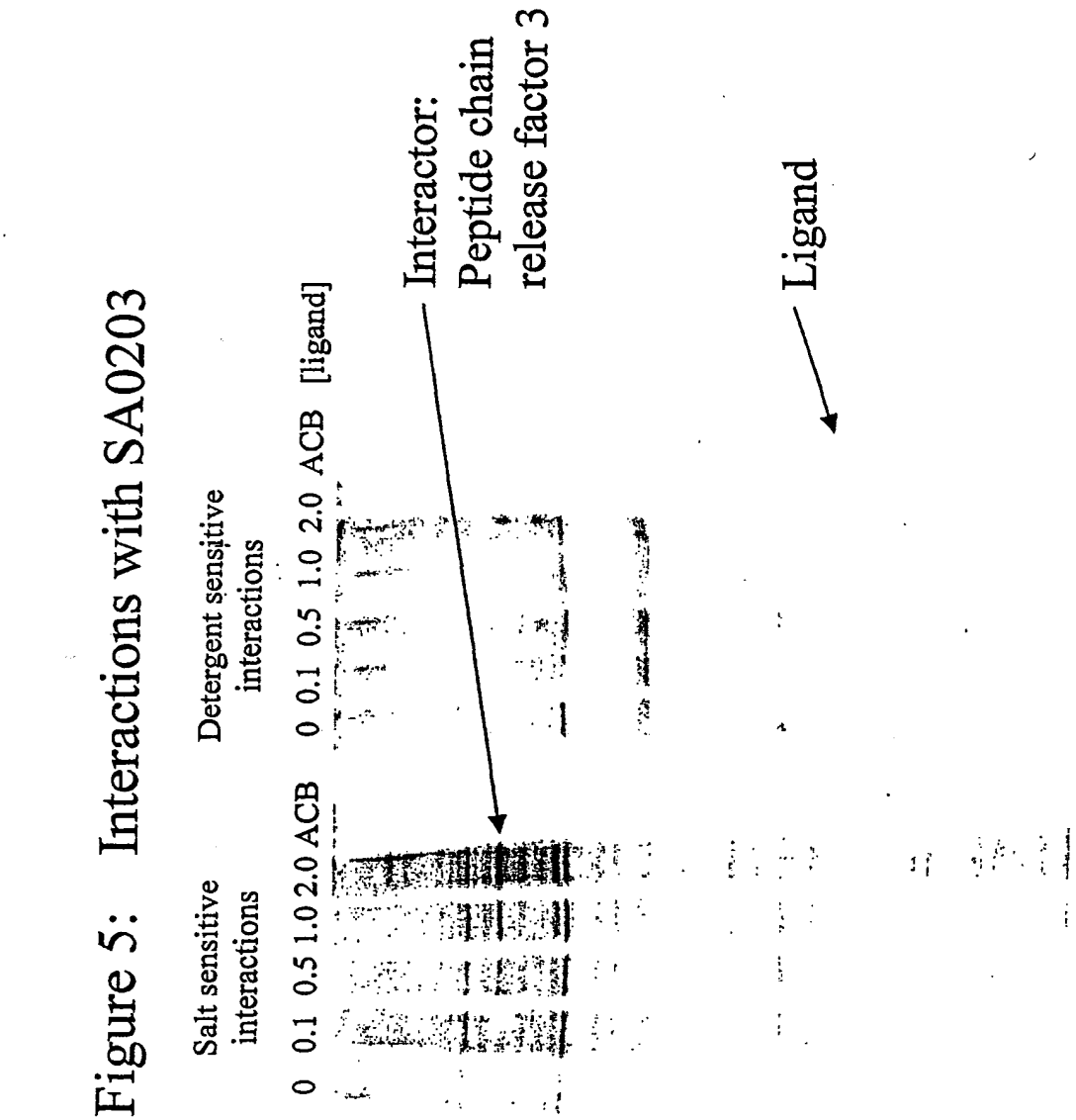


Figure 6: Interactions with SA0203

MALDI-TOF Mass Spectrum of Interactor
(peptide chain release factor 3)

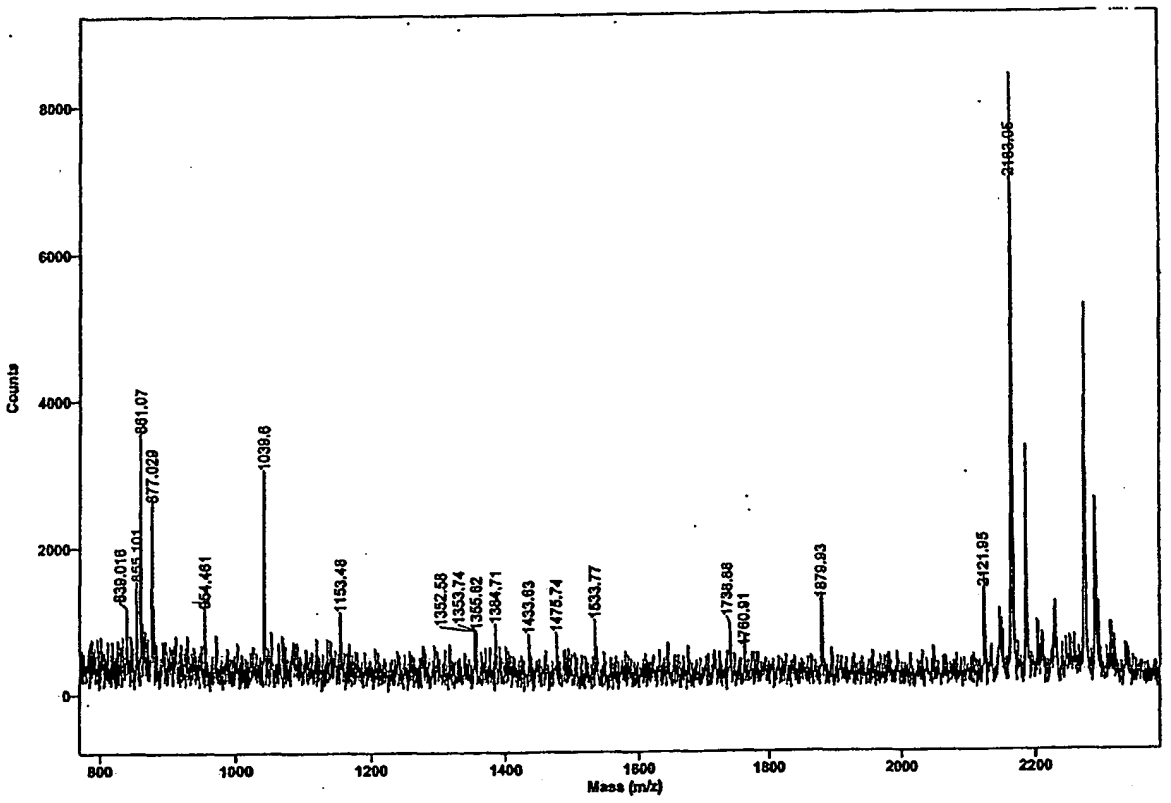


Figure 7: Interactions with SA0276

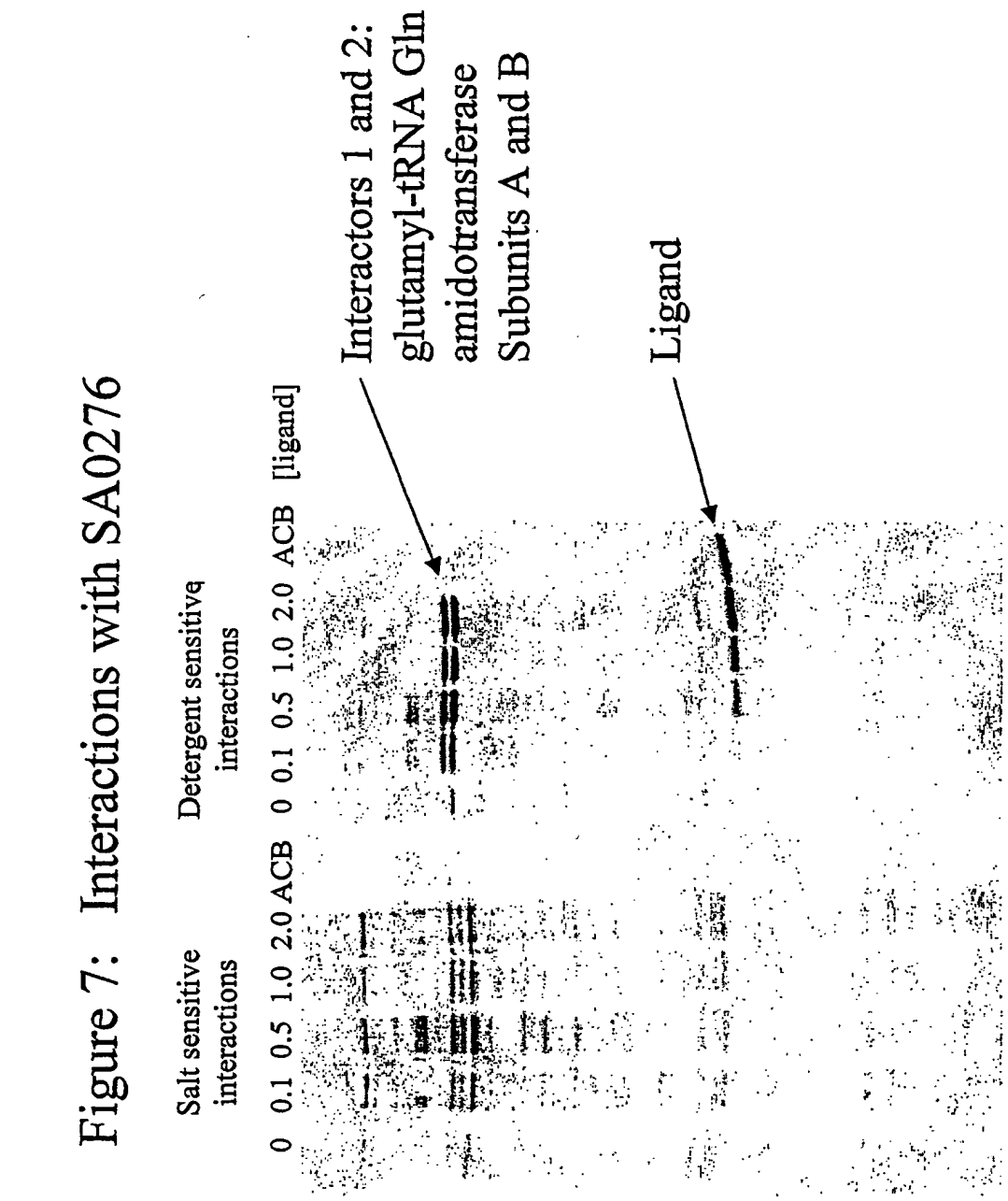


Figure 8: Interactions with SA0276

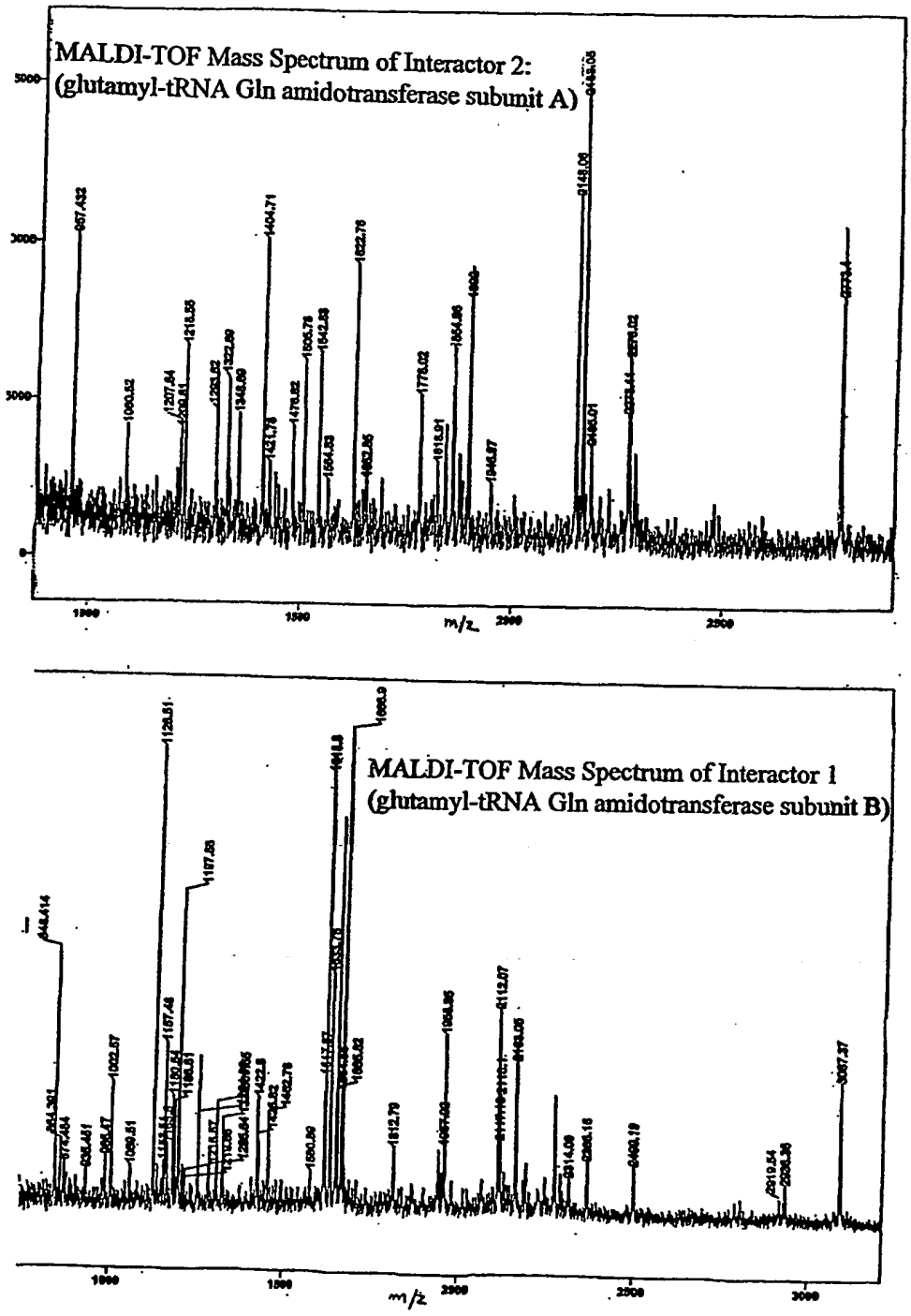


Figure 9: Interactions with SA0526

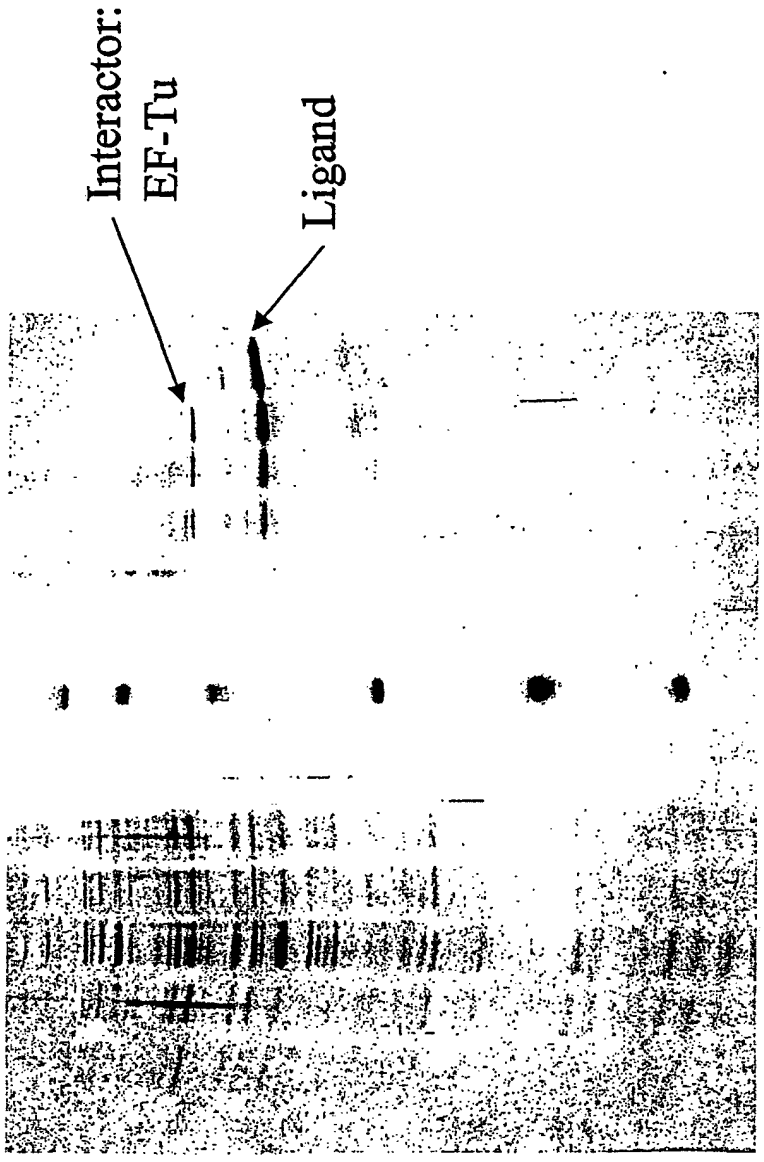


Figure 10: Interactions with SA0526

MALDI-TOF Mass Spectrum of Interactor (EF-Tu)

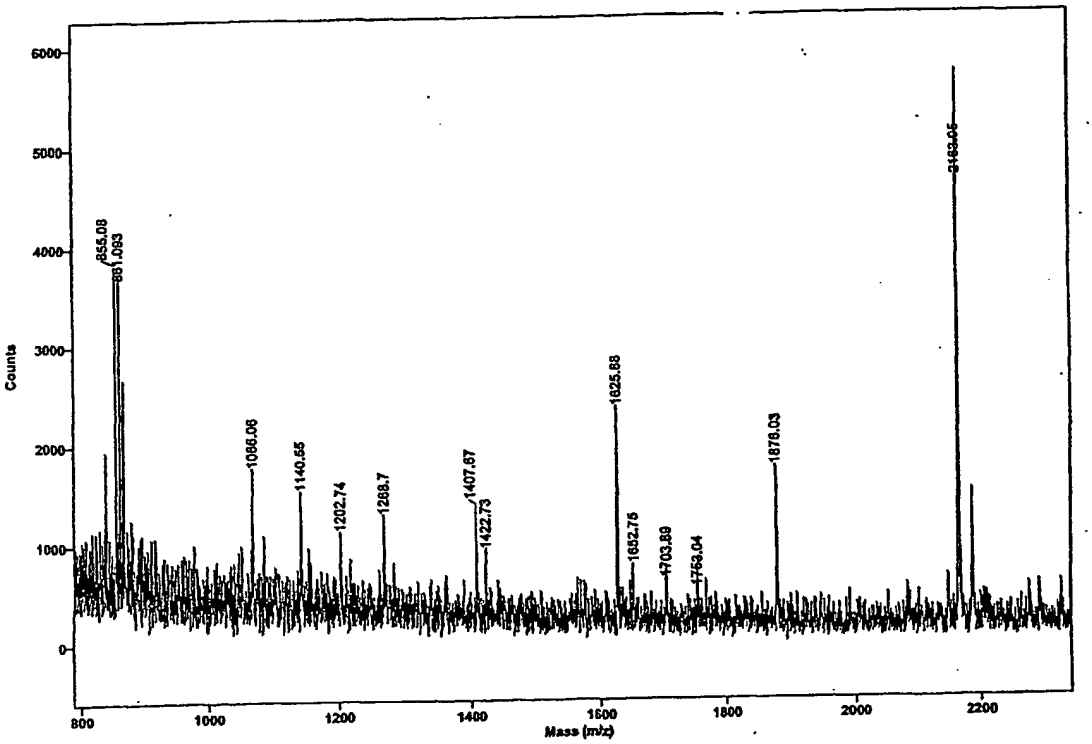


Figure 11. Interactions with SA0808

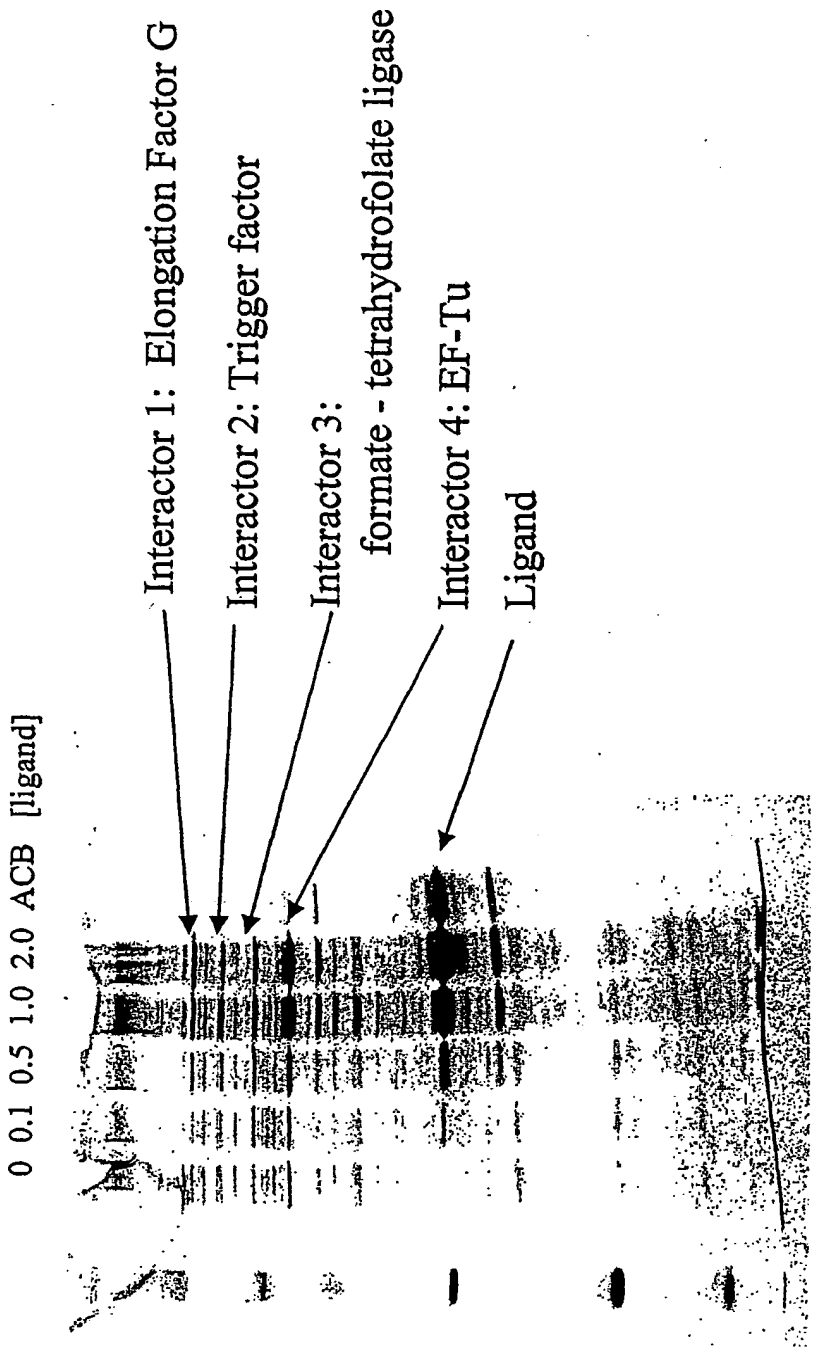
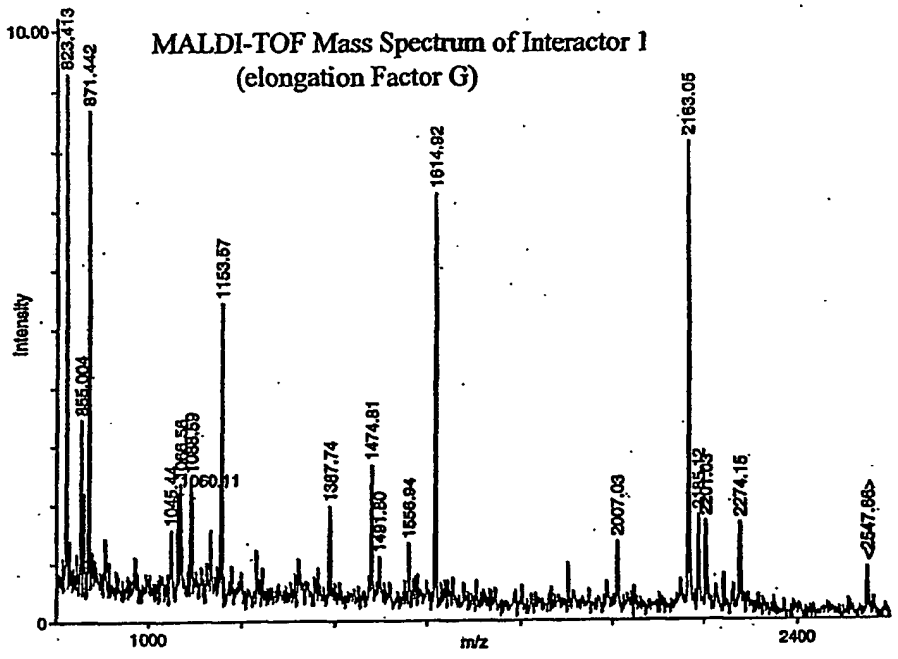


Figure 12a: Interactions with SA0808



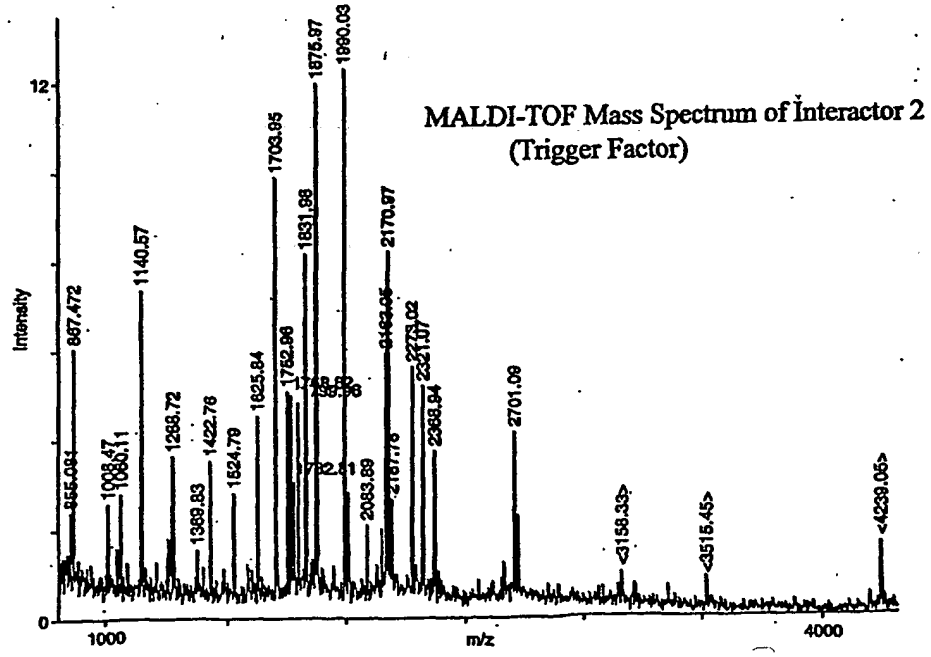
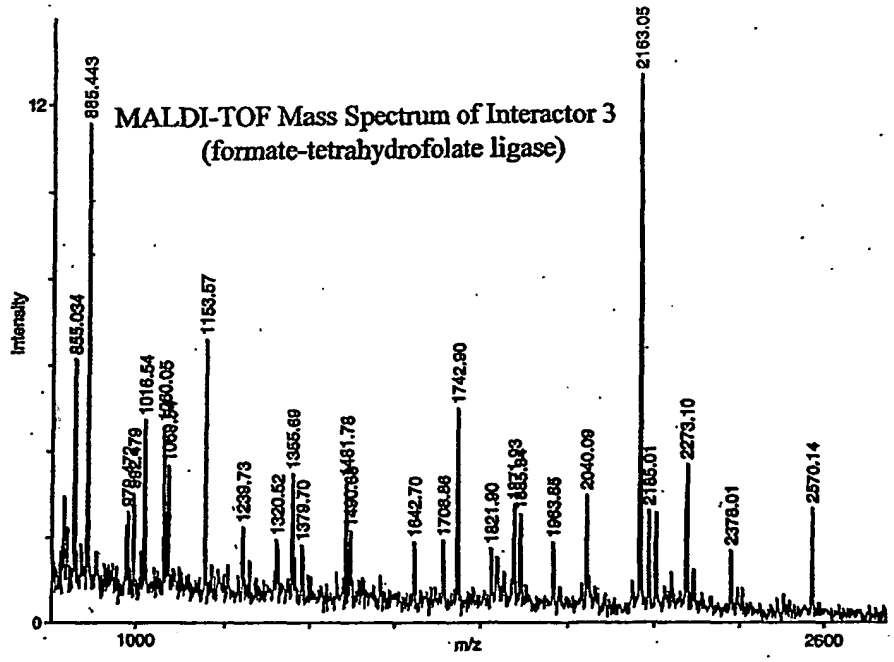


Figure 12b: Interactions with SA0808



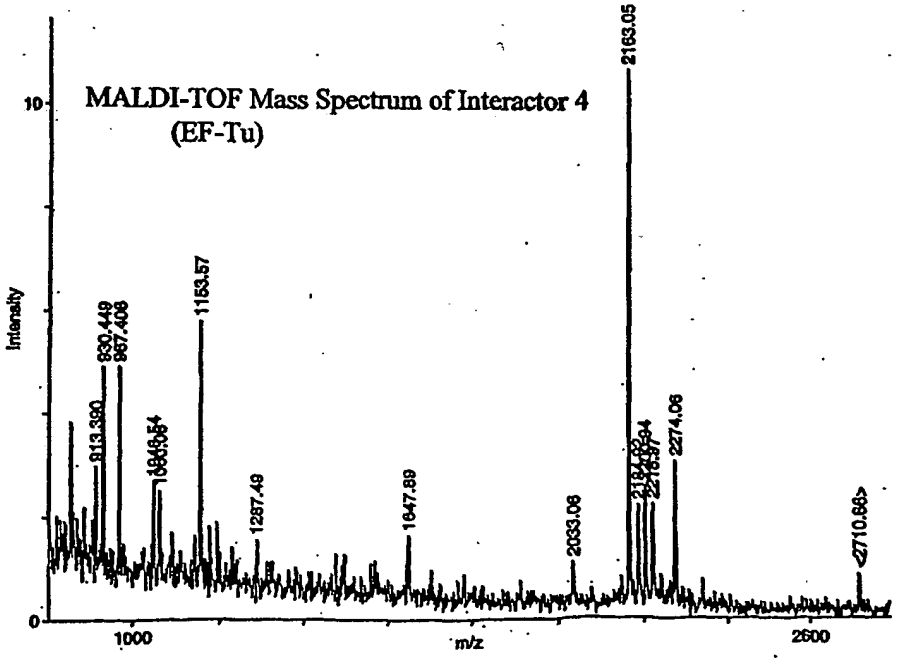


Figure 13: Interactions with SA0989

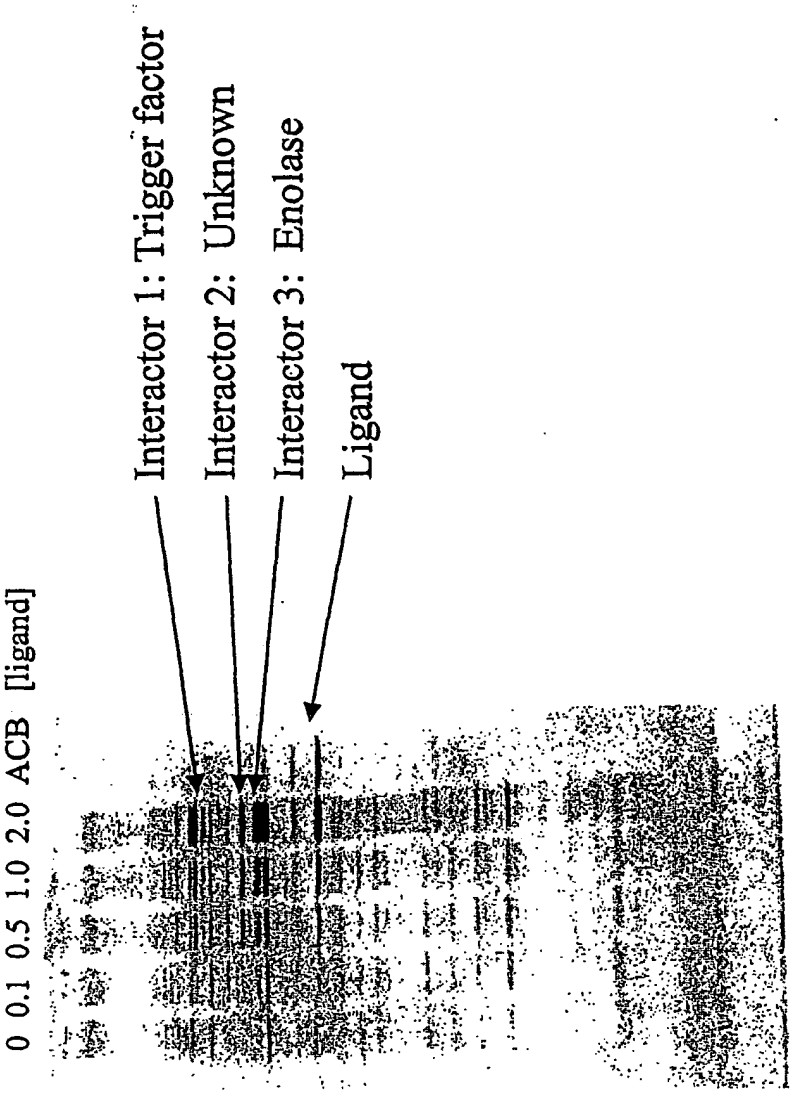
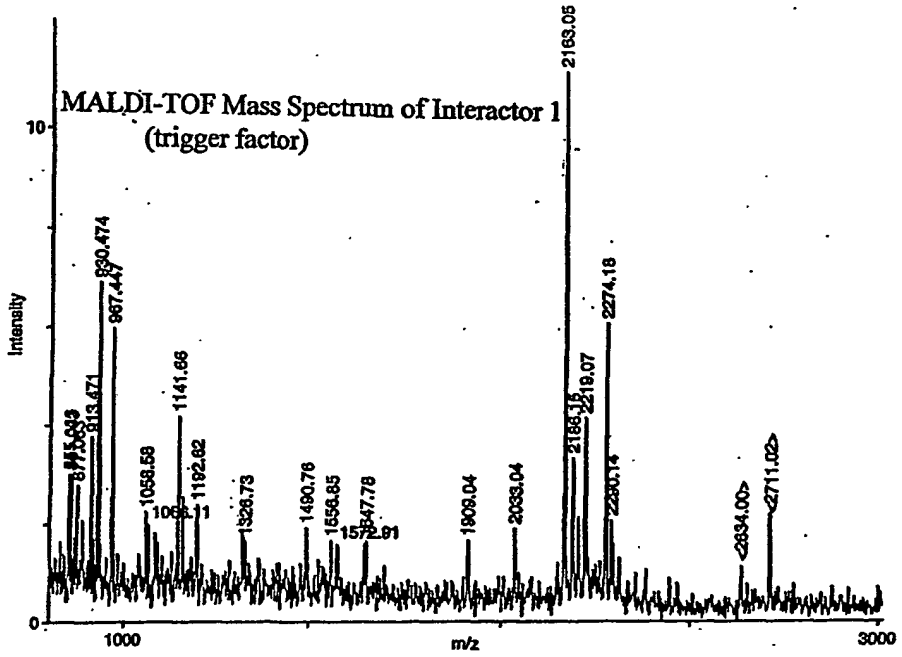


Figure 14: Interactions with SA0989



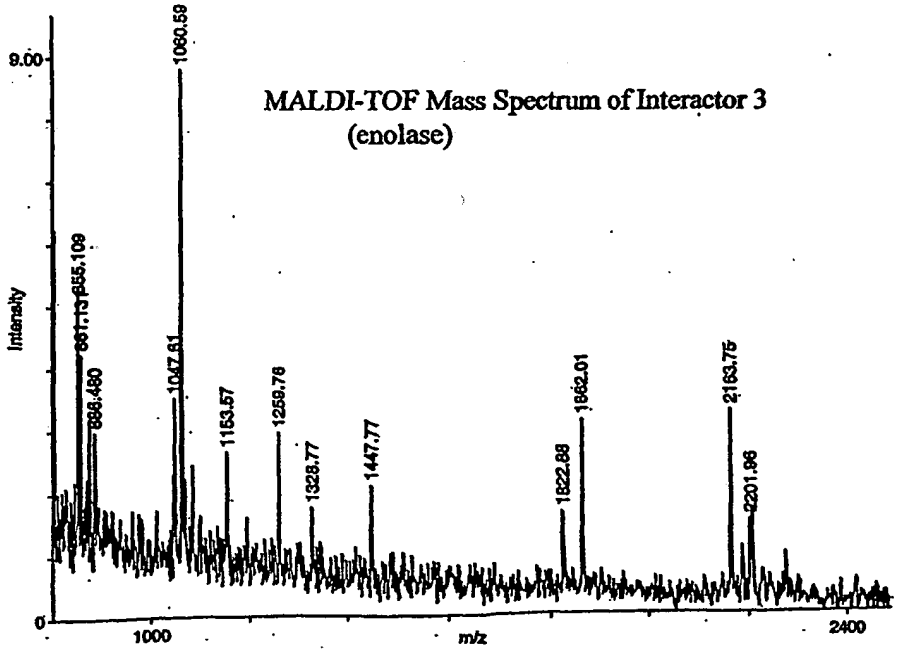


Figure 15: Interactions with SA1094

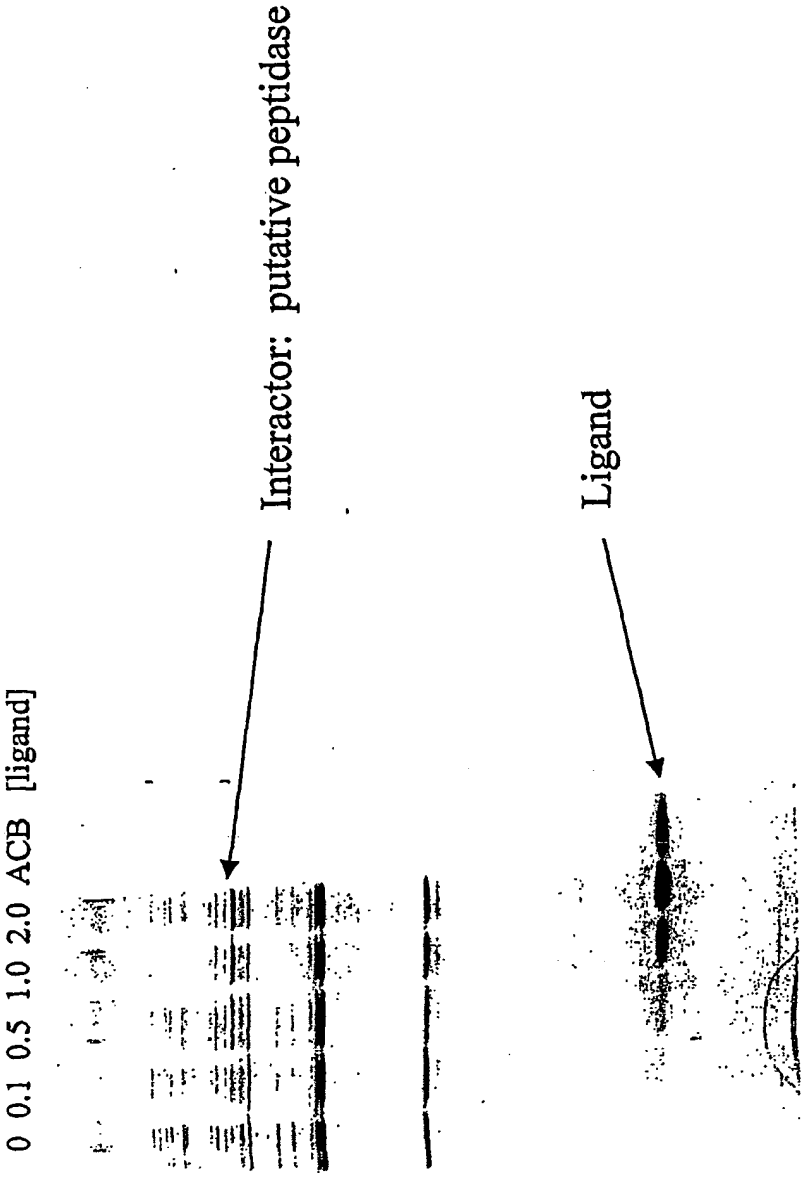


Figure 16: Interactions with SA1094

MALDI-TOF Mass Spectrum of Interactor (putative peptidase)

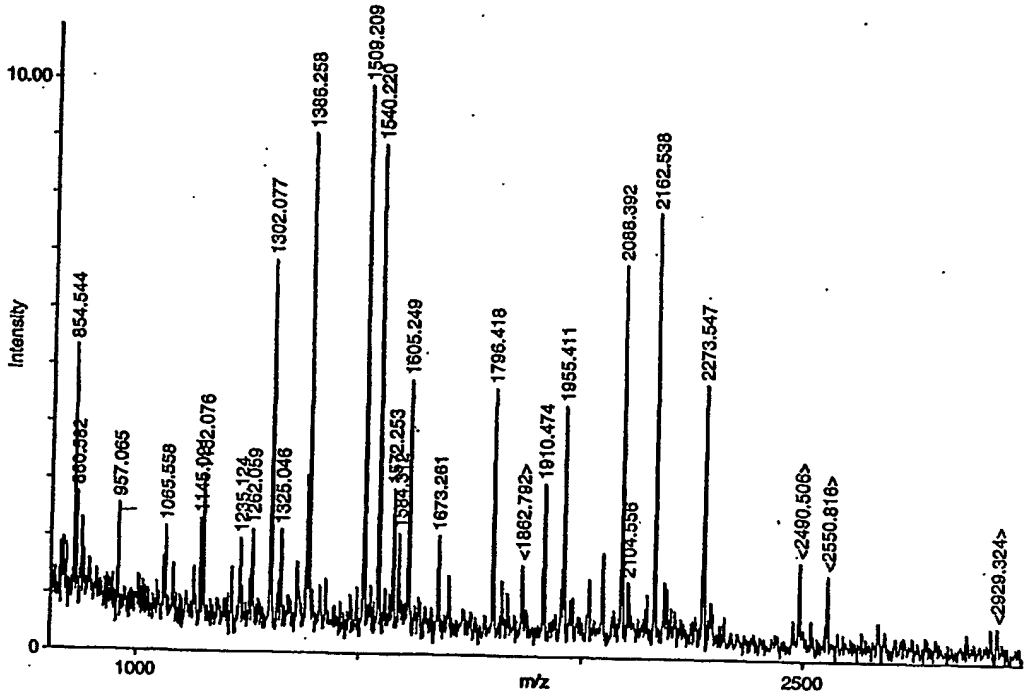


Figure17: Interactions with SA1185

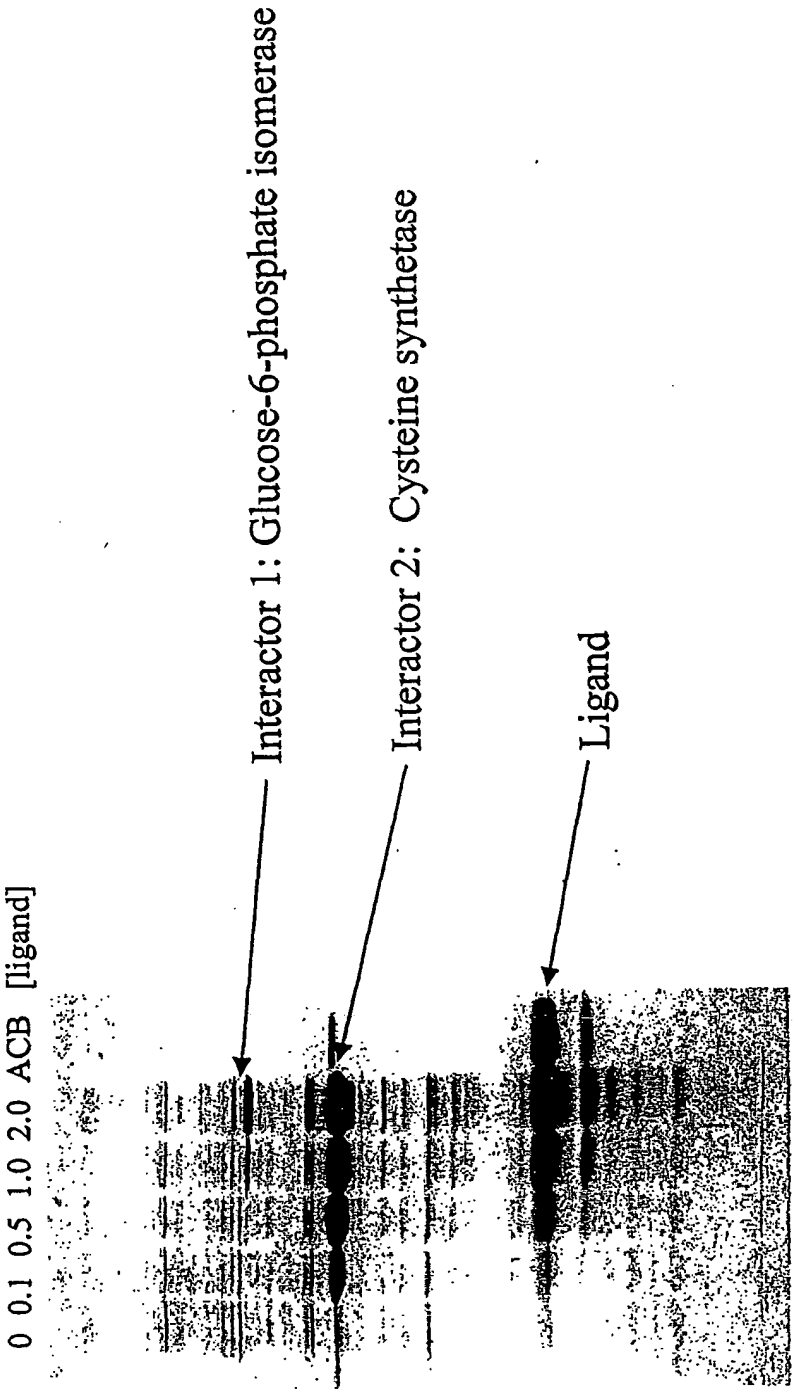


Figure 19: Interactions with SA1203

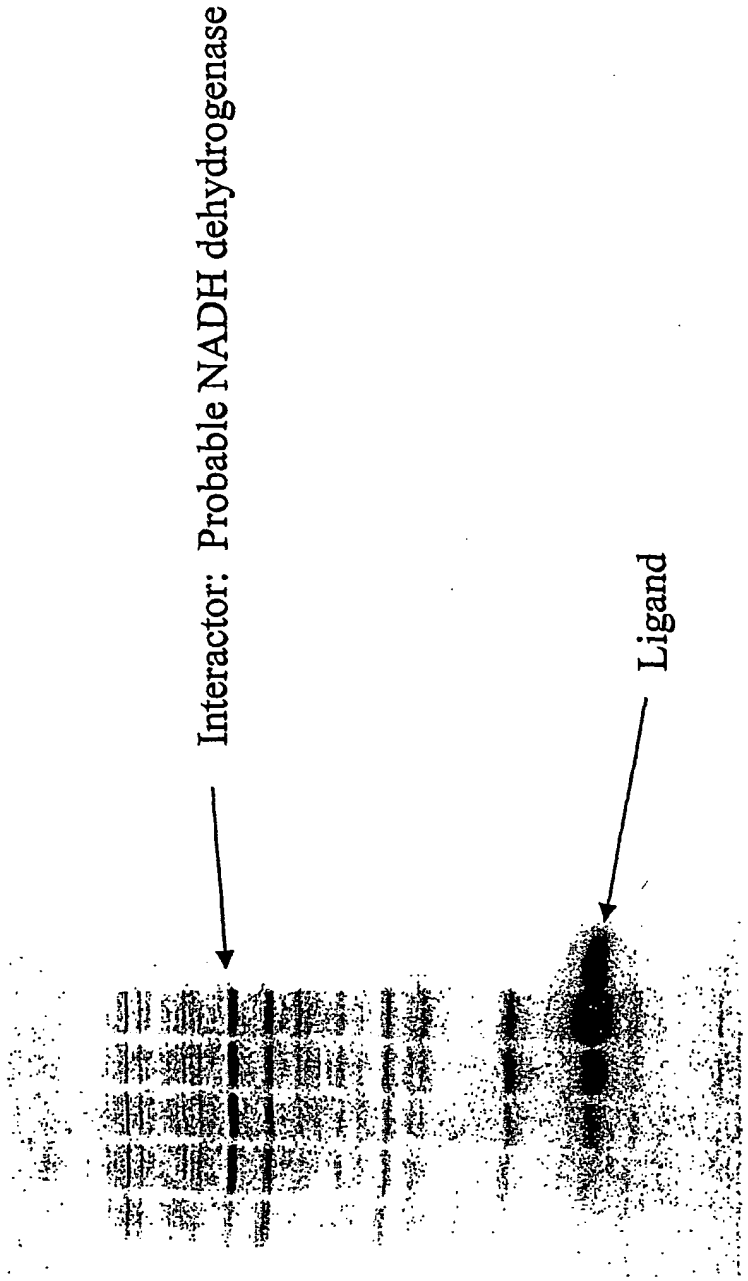
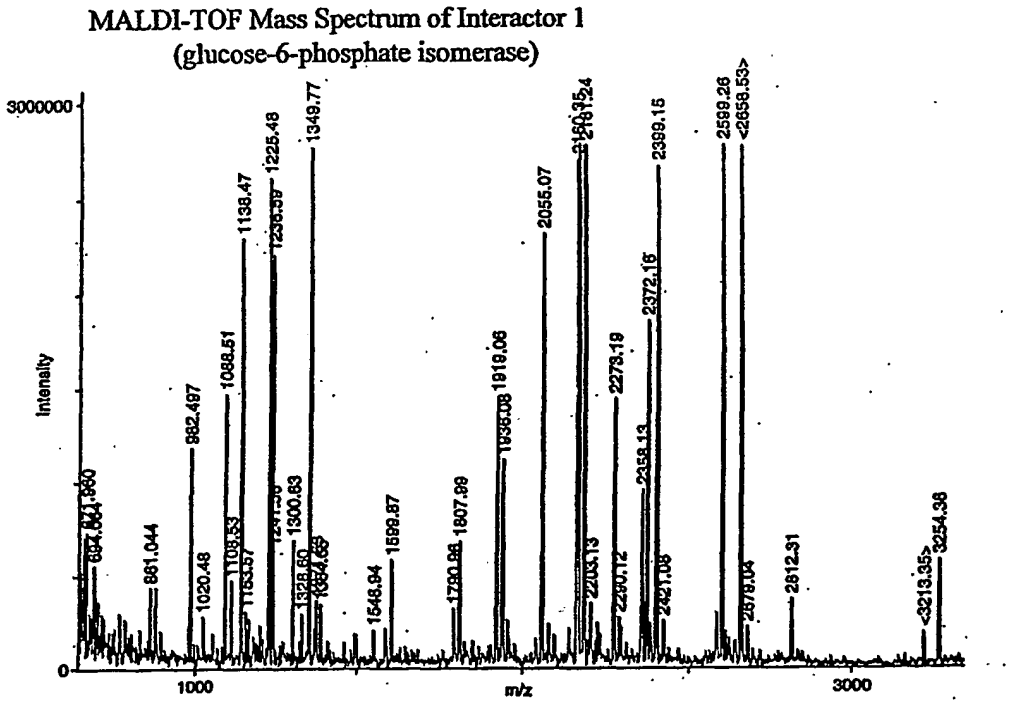


Figure 18: Interactions with SA1185



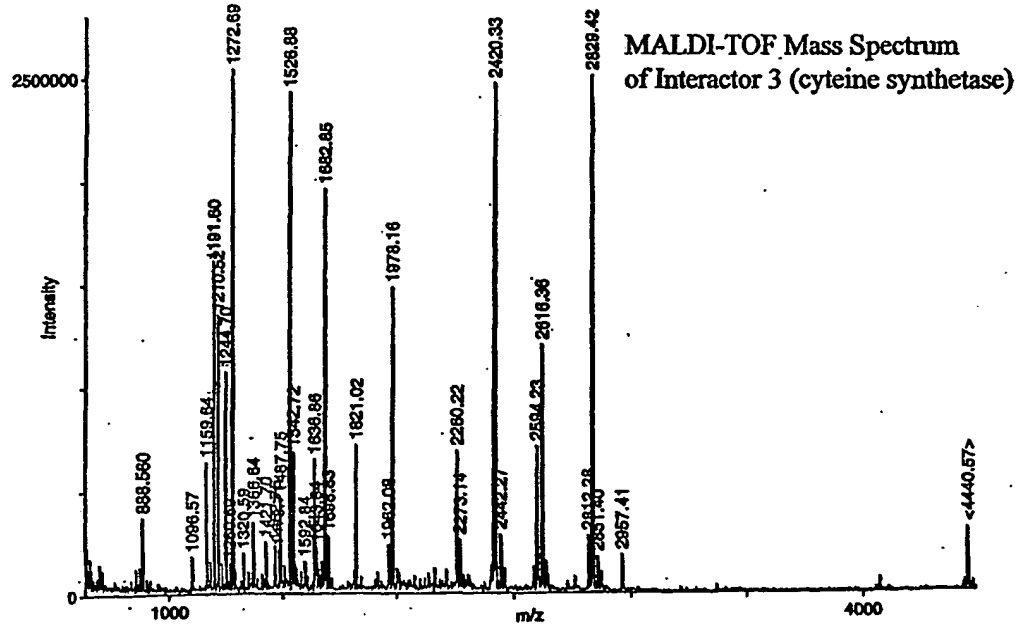


Figure 20: Interactions with SA1203

MALDI-TOF Mass Spectrum of Interactor
(probable NADH dehydrogenase)